

Comparison of Genetic Variability at Multiple Loci across the Genomes of the Major Subtypes of Kaposi's Sarcoma-Associated Herpesvirus Reveals Evidence for Recombination and for Two Distinct Types of Open Reading Frame K15 Alleles at the Right-Hand End

LYNN J. POOLE,¹ JIAN-CHAO ZONG,² DOLORES M. CIUFO,² DONALD J. ALCENDOR,¹
JENNIFER S. CANNON,² RICHARD AMBINDER,² JAN M. ORENSTEIN,³
MARVIN S. REITZ,⁴ AND GARY S. HAYWARD^{1,2*}

*Department of Oncology, The Johns Hopkins School of Medicine, Baltimore, Maryland 21231²; Department of Pharmacology and Molecular Sciences, The Johns Hopkins School of Medicine, Baltimore, Maryland 21205¹;
Department of Pathology, George Washington University Medical Center, Washington, D.C. 20037³;
and Institute of Human Virology, University of Maryland at Baltimore, Baltimore, Maryland 21201⁴*

Received 24 February 1999/Accepted 10 May 1999

Kaposi's sarcoma (KS)-associated herpesvirus or human herpesvirus 8 (HHV8) DNA is found consistently in nearly all classical, endemic, transplant, and AIDS-associated KS lesions, as well as in several AIDS-associated lymphomas. We have previously sequenced the genes for the highly variable open reading frame K1 (ORF-K1) protein from more than 60 different HHV8 samples and demonstrated that they display up to 30% amino acid variability and cluster into four very distinct evolutionary subgroups (the A, B, C, and D subtypes) that correlate with the major migratory diasporas of modern humans. Here we have extended this type of analysis to six other loci across the HHV8 genome to further evaluate overall genotype patterns and the potential for chimeric genomes. Comparison of the relatively conserved ORF26, T0.7/K12, and ORF75 gene regions at map positions 0.35, 0.85, and 0.96 revealed typical ORF-K1-linked subtype patterns, except that between 20 and 30% of the genomes analyzed proved to be either intertypic or intratypic mosaics. In addition, a 2,500-bp region found at the extreme right-hand side of the unique segment in 45 HHV8 genomes proved to be highly diverged from the 3,500-bp sequence found at this position in the other 18 HHV8 genomes examined. Furthermore, these previously uncharacterized "orphan" region sequences proved to encompass multiexon latent-state mRNAs encoding two highly diverged alleles of the novel ORF-K15 protein. The predominant (P) and minor (M) forms of HHV8 ORF-K15 are structurally related integral membrane proteins that have only 33% overall amino acid identity to one another but retain conserved likely tyrosine kinase signaling motifs and may be distant evolutionary relatives of the LMP2 latency protein of Epstein-Barr virus. The M allele of ORF-K15 is also physically linked to a distinctive M subtype of the adjacent ORF75 gene locus, and in some cases, this linkage extends as far back as the T0.7 locus also. Overall, the results suggest that an original recombination event with a related primate virus from an unknown source introduced exogenous right-hand side ORF-K15(M) sequences into an ancient M form of HHV8, followed by eventual acquisition into the subtype C lineage of the modern P-form of the HHV8 genome and subsequent additional, more recent transfers by homologous recombination events into several subtype A and B lineages as well.

Kaposi's sarcoma (KS) herpesvirus or human herpesvirus 8 (HHV8) DNA is present in virtually all tumor samples of both classical and AIDS-associated forms of KS (1a, 5), as well as in peripheral blood mononuclear cells in up to 50% of homosexual AIDS patients with KS (35). Infection without disease is also widespread in parts of Africa, the Middle East, and Mediterranean countries, with up to 60% seropositivity in Central Africa and 15% seropositivity in Italy, where classic and endemic KS rates are relatively high (11, 12, 16). Seropositivity is also very high among both KS positive and KS-negative male homosexual AIDS patients (85 and 50%, respectively) but not usually in human immunodeficiency virus (HIV)-positive in-

travenous drug user or hemophiliac AIDS patients. The seroprevalence in blood donors in the United States, the United Kingdom, and Japan may be no greater than 1% (11, 12), which also correlates with low KS incidences in renal transplant patients in those countries compared to reports of significant iatrogenic KS in patients of Middle East or Mediterranean Jewish ancestry in Toronto and in Saudi Arabia (13, 29).

HHV8 is a class gamma-2 herpesvirus that is distantly related to herpesvirus saimiri and Epstein-Barr virus (EBV) but contains several novel loci that include diverged viral homologues of exogenously acquired cellular genes encoding proteins such as interleukin-6, MIP-1A, MIP-1B, BCL-2, dihydrofolate reductase, TS, IRF, BCL-2, OX-2, FLIP, GCR, and CYC-D in place of the latent-state EBNA proteins and several other specific regulatory gene products of EBV (4, 22–27, 31). The nearly complete primary nucleotide sequences of the 190-kb double-stranded DNA molecules of two HHV8 genomes, one derived from an AIDS body cavity-based lymphoma cell line (31) and the other derived from a KS lesion

* Corresponding author. Mailing address: Department of Pharmacology & Molecular Sciences, The Johns Hopkins School of Medicine, 725 N. Wolfe St., WBSB 317, Baltimore, MD 21205. Phone: (410) 955-8684. Fax: (410) 955-8685. E-mail: ghayward@jhmi.edu.

(24), have been determined, and they differ overall by only 0.4% from each other.

Comparison of low-level sequence variability among different HHV8 samples in the open reading frame 26 (ORF26) and ORF75 gene regions originally led us to conclude that at least three distinct subtypes of HHV8 genomes occur (38). Subsequently, we found that the 289-amino-acid ORF-K1 membrane protein of HHV8 displays unusually high levels of genetic variability, resulting in four major subtypes, referred to as A, B, C, and D, that differ by 15 to 30% at the amino acid level (27, 37). Even within these ORF-K1 subtypes, amino acid differences of up to 6 to 8% among different clusters of subtype A ORF-K1 proteins and of 9 to 12% between two major branches of subtype C ORF-K1 proteins, together with small in-frame deletions, permitted the division of 46 genomes within the A-plus-C supergroup into 10 distinct variants plus several more narrowly diverged clades. Most U.S. AIDS KS samples are A1, A4, or C3 variants, whereas most classic KS cases from the Middle East, Asia, Europe, and United States are C2 variants. In contrast, samples from Africa are predominantly of the B subtype (plus occasional A5 variants) and the rare D subtype appears to be of Pacific Island origin. For example, among a total of 63 genomes examined at the ORF-K1 locus, 10 of 11 samples from central and southern Africa were subtype B; 6 of 7 from Saudi Arabia were subtype C2, C4, or C5; 7 of 9 from Taiwan represented a novel C3' clade; 6 of 7 from AIDS patients in Maryland were A1 variants, and all 3 samples from classic KS patients of native Pacific Island heritage were the only examples found of the novel D subtype.

We interpreted these results to imply that the major evolutionary branches of HHV8 correlate with the modern human population divisions that occurred via migrations from Africa first to southern Asia and Oceania 60,000 years ago and second to Europe and northern Asia some 35,000 years ago (37). The fact that cladal populations of the virus are still recognizable from paleolithic times presumably reflects both the relatively low-level distribution of the virus in many present-day populations and a relatively slow rate of horizontal transmission, which would together reduce recombination and scrambling of the differences to minimal levels. This scenario provides an appropriate evolutionary time scale for the application of some unknown but powerful biological selective pressure that generated ORF-K1 diversity. The ORF-K1 amino acid variability levels are similar to those found in immunoglobulin variable chains and in the HIV ENV protein, but we presume that unlike those situations, herpesviruses have no specialized mechanisms to rapidly generate this level of diversity. Nevertheless, it is plausible that some of the HHV8 cladal distribution patterns, especially within the hypervariable VR* loop, could reflect much more recent founder effects associated with higher horizontal infectivity rates within the AIDS epidemic. Further studies to continue to explore the patterns of HHV8 strain differences were undertaken because of the many intriguing questions raised about the molecular evolution of HHV8 genomes and the possibility that strain differences are associated with different disease states.

In the present study, we have evaluated variability at six other loci across the HHV8 genome, including three segments in the central conserved portions of the genome and three at the extreme right-hand end of the DNA molecule. Overall, the results have (i) confirmed that the divisions into three or four major subtypes can also be recognized within the more conserved portions of the genome, (ii) led to the identification of two highly diverged alternative allelic forms (referred to as predominant [P] and minor [M]) of the ORF-K15 region at the right-hand side (RHS) of the genome, and (iii) provided strong

evidence that some genomes represent mosaics derived either by recombination between major subtypes or with some unknown but related exogenous virus source. In addition, we present evidence from reverse transcription (RT)-PCR analysis that the orphan extreme RHS regions of both the P and M subtypes of HHV8 encompass complex spliced latent-state genes encoding highly diverged integral membrane proteins that are distantly related to the LMP2 protein of EBV.

MATERIALS AND METHODS

HHV8 genomic phage library and subclones. A set of phage clones derived from two partial *Sau3A* BCBL-R tumor DNA genomic libraries in λ EMBL3 and λ DASHII backgrounds were described previously (25, 38). A λ EMBL3 clone (λ B3-2) containing a 17-kb insert from position 120.6 to position 137.4 proved to encompass the RHS terminus across the boundary of the unique region into the RHS terminal tandem repeats (TTRs). A 3.3-kb *HindIII-SalI* plasmid subclone, pDJA62, encompassing all of ORF-K15(P) and the RHS TTR from a subtype A genome was generated and sequenced by shotgun M13 procedures (part of the updated GenBank ORF75/K15 file [accession no. U85269]). Two additional λ EMBL3 clones, λ E-A2 (108.3 to 120.4) and λ E-C2 (115.7 to 129.0), encompassing the T0.7 gene region were also isolated, and a 5.0-kb *EcoRI*-to-*EcoRI* plasmid subclone was generated for initial PCR sequencing of the BCBL-R T0.7 region equivalent to coordinates 115470 to 120379 of BC1 (31).

HHV8-positive DNA samples. The sources and relevant characteristics of all of the KS and primary effusion lymphoma (PEL) DNA samples used here, as well as the procedures used for extraction of archival paraffin or frozen blocks, were described previously (37, 38).

PCR amplification and sequencing primers. Direct or nested PCR products from the following regions of the HHV8 genome were generated with Bethesda Research Laboratories (BRL) *Taq* or Stratagene extend polymerase in a Techne PHC3 thermocycler set at 94°C for 1 min, 50°C for 1 min, and 72°C for 2 min over 35 cycles using 20 to 100 ng of template DNA: ORF26 (330 bp), LGH1701 [5'-(GGAT)GGATCCCTCTGACAACC-3'] and LGH1702 [5'-(ACGT)GGATCC GTGTTGTCTACG-3']; ORF75-E (854 bp), LGH2087 [5'-(CAGGTCGCTTAC TATTCTG-3') and LGH1704 [5'-(GTAC)GGATCCACGGAGCATAC-3'], as well as LGH1984 [5'-(CTAG)AGATCTGTTAGTCCGGAG-3'] and LGH2000 [5'-(GGAACAGGGTGTCTGTG-3')]; T0.7 (646 bp), LGH2076 [5'-(GCTGCAAT GTACTGCCATG-3') and LGH2075 [5'-(CTCCATCCCAATGCATGGA-3')]; ORF72 (vCYC-D) (594 bp), LGH2045 [5'-(GTAGAACGGAACATCGCA-3') and LGH2046 [5'-(GATTGGTATTGGACCTTTC-3')]; ORF-K1 (1,065 bp), LGH2089 [5'-(GTTCTGCCAGGCATAGTC-3') and LGH2088 [5'-(AATAAGT ATCCGACCTCAT-3')]; K14.1(P) (362 bp), LGH2079 [5'-(GAGATCACTCTC CAACCCAC-3') and LGH2033 [5'-(GGAGTGCTTCCGTATAG-3')]; K14.1(M) (450 bp), LGH2079 [5'-(GAGATCACTCTCCAACCAC-3') and LGH2506 [5'-(CACAGTCACCTATGCTAG-3')]; K15(P) (580 bp), LGH2476 [5'-(GCAGTGT TTATTAACGTC-3') and LGH2477 [5'-(CAAAACCCATTACTTC-3')]; K15 (M) (370 bp), LGH2473 [5'-(CATGCAGCGAGCTTGAGA-3') and LGH2474 [5'-(CTTTGAGTACTGTTGTG-3')]; U/TTR (850 bp), LGH2098, [5'-(AAGAT ATAGACCCACATAC-3') and LGH2097 [5'-(CACGTAGCAAGCACTGAG-3')]. The underlined sequences represent added restriction sites for cloning purposes.

The PCR DNA products were fractionated on 1% agarose gels (GIBCO BRL ultrapure catalog no. 15510-027), stained with ethidium bromide, and photographed under longwave UV light. Isolated bands were cut out from the gel and extracted with QiaexII glass milk suspension (Qiagen Gel Extraction Kit catalog no. 20021) and subjected to manual ³²P-labelled dideoxynucleotide double-stranded cycle DNA sequencing (GIBCO BRL catalog no. 18196) using both left-hand side (LHS) and RHS external, and in some cases also internal, primers. Virtually all of this sequence data were generated on both complementary strands, and many analyses included redundant overlapping runs or duplicative sequencing on multiple independent PCR-amplified products. Whenever possible, direct PCR products were used as sequencing templates in preference to nested amplification products.

Cloning and sequencing of RT-PCR products. Total RNA was isolated from uninduced HBL6 or BCBL-1 cells by using Trizol reagent (GIBCO BRL). For each cell line, RNA samples (3 μ g each) were either DNase treated and subjected to RT-PCR or not DNase treated but subjected to RT-PCR or DNase treated and subjected to DNA PCR. DNase treatment involved incubation with 10 U of RNase-free DNase I (Boehringer Mannheim) for 4 h at 37°C. The samples were then extracted with phenol-chloroform, and the RNA was precipitated in 2.5 volumes of ethanol and 1/10 volume of 3 M sodium acetate (pH 5.2). Samples were resuspended in diethylpyrocarbonate-treated distilled H₂O, and reverse transcription was carried out at 50°C for 50 min by using Superscript II (GIBCO BRL) and an oligo(dT) primer (Boehringer Mannheim). Subsequent DNA amplification was performed by PCR for 45 cycles using *Taq* DNA polymerase (Promega) with the following 5' and 3' primers designed to encompass the complete ORF-K15 coding region for both the P and M forms: HBL6(M) N terminus (LGH3129), 5'-CTAGGAATTTCATGAATTACAAAAATACCTGT GGGG-3'; HBL6(M) C terminus (LGH3132), 5'-CTAGGAATTCGTCGGTG GGAACAAAAAC-3'; BCBL-1(P) N terminus (LGH3110), 5'-CTAGGGATCCA

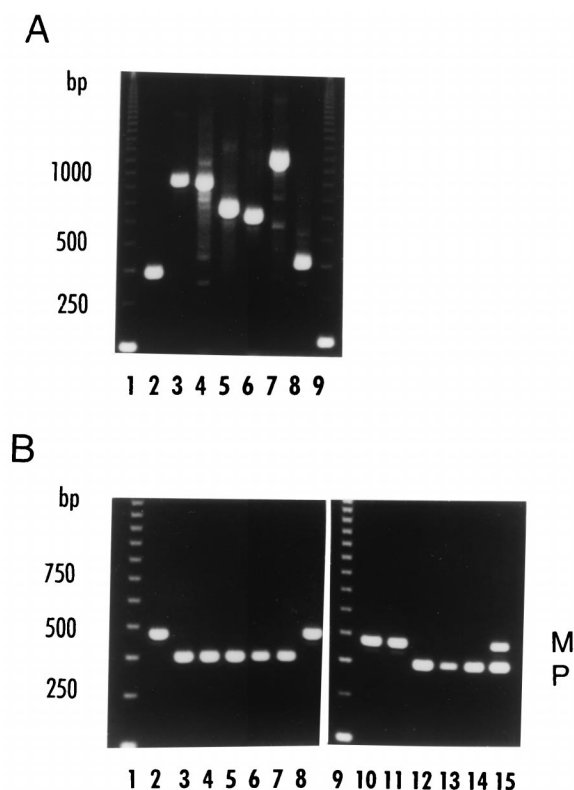


FIG. 2. Representative examples of PCR DNA products involved in these studies. (A) Photograph of ethidium bromide-stained DNA bands in an agarose gel displaying a complete set of directly amplified PCR products from across the HHV8 genome present in the JSC1 cell line for use in DNA sequence analysis. Lanes: 1 and 9, 123-bp oligomer size marker ladder; 2, ORF26 primers (LGH1701 and LGH1702); 3, ORF75-E (LGH2087 and LGH1704); 4, U/TTR (LGH2097 and LGH2098); 5, T0.7 (LGH2076 and LGH2075); 6, CYC-D (LGH2044 and LGH2045); 7, ORF-K1 (LGH2089 and LGH2088); 8, ORF-K14.1 (triple-primer reaction of LGH2079, LGH2033; and LGH2506). (B) Photographs of ethidium bromide-stained RHS junction region PCR DNA products (K14.1) from various P (362 bp) and M (450 bp) allele genomes. Lanes 1 and 9 contained a 123-bp oligomer size marker ladder. The KS and PEL DNA samples used included the following: lanes 2 and 10, HBL6 (M); lane 3, BC2 (P); lane 4, BC3 (P); lane 5, AKS1 (P); lane 6, BCP1 (P); lane 7, BCBL-B (P); lane 8, ASM72 (M); lane 11, ASM76 (M); lane 12, BCBL1 (P); lane 13, EKS1 (P); lane 14, C282 (P); lane 15, a mixture of BCLB1 and HBL6 DNAs. The triple-primer combination of LGH2079 (both) plus LGH2033 (P) and LGH2506 (M) was used in standard direct PCRs.

431KAP) revealed that although subtypes A and B continued to be nearly identical upstream of the ORF75 coding region, both subtype C patterns proved to diverge toward a nearly totally different sequence beginning within 300 bp of the N terminus of the ORF75 coding region. To analyze the remainder of the RHS unique region from our subtype A genome, we first mapped and subcloned a 16-kb genomic region from the extreme RHS unique-segment terminus of BCBL-R DNA in a selected phage lambda clone (λ B3-2) that proved to encompass both the ORF75 region and the adjacent G-plus-C-rich TTR sequences (Fig. 1B). M13 sequencing on both strands of a 3.3-kb *HindIII*-to-*SalI* subclone containing the junction fragment (in plasmid pDJA62) revealed the presence of a 2,354-bp unique sequence between the ORF75 initiator codon and the beginning of the TTR region. The junction with the RHS unique region begins at position 653 of the published TTR sequence of HHV8(BC1) (GenBank accession no. U75699) (31) and then crosses into the next TTR unit and ends at position 551, representing the *Sau3A* site at the end of the RHS of λ B3-2 (Fig. 1B). As judged by the finding of identical-size PCR-amplified DNA products for a region spanning the

unique-region-TTR junction from the original BCBL-R tumor DNA, as well as in both the λ B3-2 lambda clone and the M13-sequenced pDJA63 plasmid DNA (data not shown), we conclude that this sequence represents the complete undelimited RHS unique region from a prototype subtype A genome (GenBank accession no. U85269). Subsequent PCR sequencing of BCBL-1 DNA in this region also showed it to be almost identical to that of BCBL-R.

In contrast, when the complete genome sequence data from Russo et al. (31) for BC1 became available, it proved to have a 3,500-bp unique sequence upstream of the ORF75 gene in a region equivalent to that to the right of where the subtype A and C patterns begin to diverge, together with an unclonable G+A-rich repeat segment whose size is unknown, before reaching the first copy of the 801-bp TTR sequence at the extreme RHS of the genome (Fig. 1B). Surprisingly, a more than 2,000-bp unique region sequence to the right of ORF75 in our BCBL-R sequence showed no significant nucleotide homology to the RHS (or any other sequences) in the published complete HHV8(BC1) genome (GenBank accession no. U75698). Data for this region are missing in the nearly complete HHV8 KS sequence presented by Neipel et al. (24) (GenBank accession no. U93872), but subsequent PCR sequence data for the equivalent regions from our prototype subtype C (ASM72) genome and HBL6 confirmed that they were both also virtually identical to the BC1 pattern across the ORF75-to-ORF-K15 junctions.

An illustration of the pattern of nucleotide differences across the junction between the common and divergent regions on the RHS between our prototype subtype A sequence in BCBL-R (and BCBL1) compared to the subtype C sequence in HBL6 (and BC1) is given in Fig. 3. The differences begin within the proximal N-terminal domain of the leftward-oriented ORF75 coding region, proceed across the presumed promoter region for ORF75, and end with nearly total divergence beginning 350 bp upstream of the ORF75 ATG codon. Interestingly, both types of sequence encompass a small potential rightward-oriented coding region of 79 amino acids (designated the ORF-K14.1 gene here), which lies precisely within the 240 bp between the N terminus of ORF75 and the C terminus of ORF-K15 and spans the beginning of the divergent region. The presumed consensus TATAAA motifs for ORF-K14.1 represent the beginning of the junction region sequences shown in Fig. 3. Residual patchy nucleotide homology beyond the ORF-K14.1 coding region represents several short, conserved amino acid motifs (highlighted) within the extreme C terminus of the leftward-oriented coding region from exon 8 of ORF-K15 (see below). The two DNA sequences become totally nonhomologous beyond this point as they progress further into the ORF-K15 genes.

Many subtype A, B, and C HHV8 genomes contain the alternative M form of the RHS junction region. To evaluate the relative abundance of each of the two types of RHS among HHV8 genomes, we generated a three-primer PCR set containing a single primer within the LHS common region at position +42 in ORF75 and two alternative RHS primers in the divergent regions either at position -320 of BCBL-R (generating a 362-bp product) or at position -408 of HBL6/BC1 (giving a 450-bp product including genome positions 134399 to 134848). Representative examples of the results are shown in Fig. 2B, and the positions of the primers are indicated in Fig. 3. Analysis of all 63 different HHV8 DNA samples for ORF-K14.1 PCR product size variations revealed 45 of the BCBL-R/BCBL1 type and 18 of the HBL6/BC1/ASM72 type. The latter were associated with more than half (13 of 24) of the genomes with subtype C ORF-K1 genes plus 3 of 19 subtype A

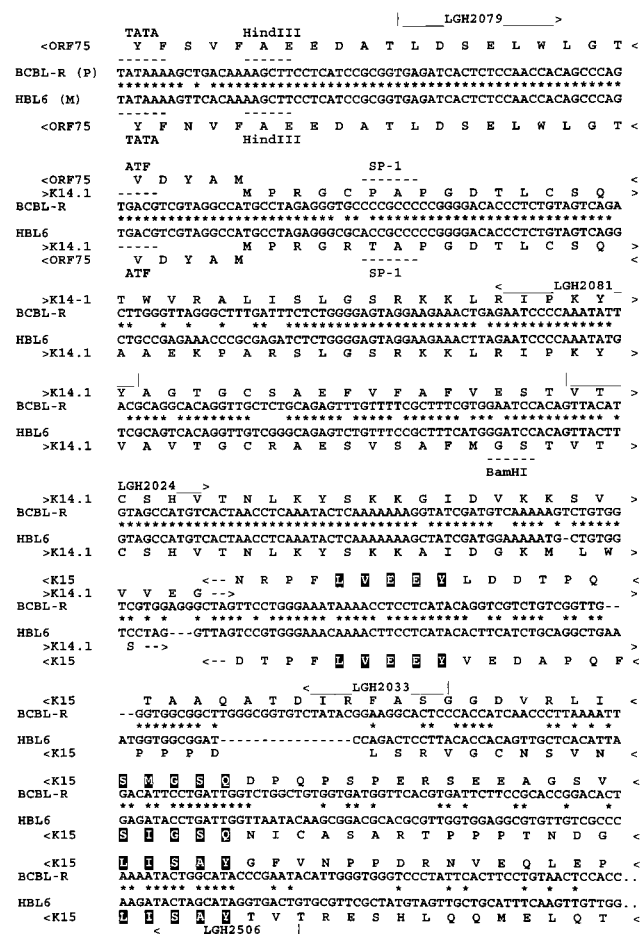


FIG. 3. Comparison of DNA sequences across the divergent junctions between the leftward-oriented ORF75 and ORF-K15 gene regions in both the P and M subtypes of HHV8. The hypothetical rightward-oriented ORF-K14.1 genes encompassed by this region are also shown. Potential transcriptional control elements and the predicted amino acids encoded by the extreme N terminus of ORF75, the extreme C terminus of ORF-K15, and the intervening ORF-K14.1 region are shown. Nucleotide identities between the P allele in BCBL-R (top line) and the M allele in HBL6 (lower line) are indicated by asterisks, and missing bases are represented by hyphens. The positions of the PCR primers, LGH2079, LGH2081, and LGH2506, that were used to discriminate between the P and M allele forms of ORF-K14.1 in the triple-primer PCR (Fig. 2B) are indicated. The nucleotide sequences shown correspond to genomic positions 1268 to 1803 in the P allele (BCBL-R; GenBank accession no. U85269) and 134365 to 134881 in the M allele (BC1; GenBank accession no. U75698).

genes and 2 of the 3 American subtype B genomes but not with any of the 12 African subtype B genomes or the 3 Pacific subtype D genomes. Therefore, the two versions of the RHS ORF-K14.1 region appear to be only partially linked to the LHS ORF-K1 subtype patterns. To avoid confusion and to clearly distinguish them from the subtype C nomenclature used for ORF-K1, we have altered the previously introduced subtype C terminology for the ORF75 regions found in ASM72 and HBL6/BC1 (38) by referring to these alternative minor population forms of the RHS as M alleles compared to the predominant P alleles found in BCBL-R and the majority of the HHV8 genomes tested.

Importantly, all of the HHV8-positive DNA samples that we tested with the three-primer ORFK-14.1 PCR assay gave only one of either the larger (M) or smaller (P) type of product and never both (Fig. 2B). However, to confirm that the situation always reflected the presence of only one or the other region as

an alternative, rather than representing preferential amplification or deletion of one of these regions from a putative larger wild-type structure encompassing both, we designed another set of four primers representing a mixture of the ORF-K15(M) gene region (370 bp including BC1 genome coordinates 136044 to 136413) and the positionally equivalent ORF-K15(P) region (580 bp) including BCBL-R positions 3125 to 3704. Again, 10 of the 10 M allele genomes tested gave 370-bp PCR products only and 10 of the 10 P allele genomes tested gave 580-bp PCR products only (data not shown).

U.S. HHV8 samples with the HBL6/BC1-like M type of RHS included both of our ORF-K1 C1 variant genomes (ASM72 and BKS13), one with a C3 variant ORF-K1 sequence (BKS12), two A2 variants (HBL6/BC1 and WKS1), one A1 variant (BKS16), and both subtype B genomes from Florida (OKS7 and OKS8). In addition, all seven of the genomes with clade C3' ORF-K1 genes and the one C2 variant (TKS11) from Taiwan were of this type, together with two of seven tested renal transplant KS samples from Saudi Arabia (SKS3 and SKS9), although many other subtype C2 and C3 genomes were not. For a summary of the distribution of P and M allele RHS ORF-K15 patterns relative to the assigned ORF-K1 subtypes, see the K14.1 and K15 columns in Fig. 9 (see also the asterisks in Fig. 6, 7, and 8). Genomes with subtype M RHS alleles are also designated by solid-circle symbols in the summary of ORF-K1 data given in Fig. 4 of reference 37.

Finally, to determine whether there is any significant variation among either the P or M alleles within the divergent ORF-K14.1 block, the products of the triple-primer PCRs were sequenced for 12 different HHV8 samples (for a summary, see Fig. 9). However, among the genomes with the P allele at the RHS terminus, all six of the A and C variants tested were identical over a total of 362 bp whereas the B version (431KAP) showed only a 1-bp change. Similarly, all five of the M allele variants tested from both ORF-K1 subtype A and C genomes were identical to one another across the equivalent 450-bp divergent region.

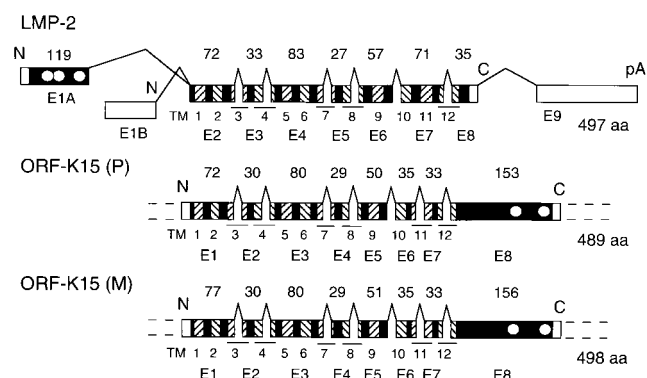


FIG. 4. The structures and intron-exon boundaries of the ORF-K15(P) and ORF-K15(M) proteins and mRNAs are remarkably similar and resemble those of the EBV LMP2 protein. Overall spliced structure and organization of coding exons of sequenced RT-PCR products from uninduced latent-state mRNAs from the BCBL-1 (P) and HBL6 (M) PEL cell lines are shown. Numbers and solid bars show the relative size and amino acid (aa) content of each coding exon (E1 to E8) from ORF-K15 (P) and ORF-K15 (M) mRNAs compared to the organization and size of exons (E1A to E9 or E1B or E9) from EBV LMP2A and LMP2B mRNAs. Seven additional amino acids are created across the splice junctions in both ORF-K15 mRNAs. Hatched and solid bars indicate coding regions, and hatched regions denote the 12 TM domains. Open bars indicate noncoding regions and exons. Circles represent conserved consensus YXXL-type SH2-binding tyrosine kinase interaction motifs. pA, polyadenylation signal.

TABLE 1. Comparison of ORF-K15(P) and ORF-K15(M) splicing signals^a

Gene (form) and domain	Defining motif	Position ^b	Size (bp) of:	
			Exon	Intron
BCBL-R (P)				
Exon 1/initiator	<u>ATG</u>	3652	(217)	
Intron 1	G'GTAAGT	3435		109
Exon 2	TTTGTTTTTATAG'	3326	93	
Intron 2	G'GTAAGG	3233		83
Exon 3	TTGTATTTTATAG'	3150	243	
Intron 3	G'GTAAGT	2907		87
Exon 4	ATCTTTTTTATAG'	2821	90	
Intron 4	G'GTACAG	2730		77
Exon 5	ACATTTTTTGTAG'	2653	153	
Intron 5	G'GTTTGT	2500		82
Exon 6	ATGTATAAACAG'	2419	108	
Intron 6	G'GTAGGT	2310		82
Exon 7	CTCTATTTTITAG'	2228	102	
Intron 7	A'GTAAGT	2126		85
Exon 8	TTTTAAAATTAG'	2041	(461)	
Terminator	<u>TAG</u>	1581		
BC1 (M)				
Exon 1/initiator	<u>ATG</u>	136771	(232)	
Intron 1	G'GTAAGT	136540		85
Exon 2	TCTTTTGTTTITAG'	136455	93	
Intron 2	G'GTAAGT	136362		96
Exon 3	ATGTGTTTTTCAG'	136266	243	
Intron 3	G'GTAAGT	136023		85
Exon 4	TTTACTACTACAG'	135938	90	
Intron 4	G'GTATAT	135848		87
Exon 5	AACTTTATTCAG'	135761	156	
Intron 5	G'GTATGT	135605		94
Exon 6	TTTTATTTACAG'	135511	108	
Intron 6	G'GTAAGG	135403		78
Exon 7	TTTATTTCCCTTAG'	135325	102	
Intron 7	A'GTGAGT	135223		80
Exon 8	TTTTTCTTATTAG'	135143	(470)	
Terminator	<u>TAA</u>	134674		

^a Predicted splicing patterns and intron and exon sizes and structures of the ORF-K15(P) and ORF-K15(M) genes of HHV8 compared to those for LMP2 in EBV (17). The data are based on DNA sequence results from cloned RT-PCR products of single intact mRNA species encompassing all seven introns obtained from latent-state mRNAs from the BCBL1 (P) and HBL6 (M) PEL cell lines. The underlined sequences represent key consensus positions. Parentheses indicate coding positions only.

^b BCBL-R, GenBank accession no. U85269; BC1, GenBank accession no. U75698.

Evaluation of the coding capacity of the two ORF-K15 regions. In their original analysis of HHV8(BC1) coding regions, Russo et al. (31) identified only a single short, leftward-oriented coding region of 100 amino acids (termed ORF-K15) within the 3,500-bp region between ORF75 and the RHS TTR. Although our DNA sequence in the equivalent 2,500-bp block in HHV8(BCBL-R) revealed no significant nucleotide homology, it displayed a very similar overall distribution of GC-rich and AT-rich patterns, as well as a small ORF-K15-like region, but with no associated ATG codon. Furthermore, both sequences also contained numerous potential consensus splice signals located on one strand only (leftward orientation) together with a second potential larger ORF without ATG motifs just upstream of ORF75. Most strikingly, there were also several very small ORFs with potentially conserved amino acids, as well as numerous potential hydrophobic transmembrane (TM) domain-like domains present at apparently colinear loci in the leftward orientation (27).

Subsequent extensive visual comparison of both unique sequences revealed that each region could potentially produce complex spliced mRNAs (Fig. 4) consisting of eight almost exactly matching coding exons spliced in a linear manner to create either a 489-amino-acid (P) or a 498-amino-acid (M)

protein. These cDNAs are predicted to encode large intact integral membrane proteins consisting of 12 matching hydrophobic TM-spanning domains and a relatively large C-terminal cytoplasmic tail (Fig. 4). Several alternative spliced patterns appear plausible as well, especially toward the N terminus of the P form. Overall, this prototype structure is remarkably similar to that of the 497-amino-acid LMP2 latency protein of EBV, which also contains 12 TM domains encoded in multiple short exons and represents an exact positional analogue of ORF-K15 upstream of the ORF75 homologue (BNRF1) located at the equivalent end of that genome.

The existence of latent-state HHV8 mRNAs containing each of the seven predicted splice junctions listed in Table 1 was confirmed by sequencing of the most abundant RT-PCR products from uninduced mRNA obtained from both the BCBL1 (P) and HBL6 (M) cell lines. In both cases, all seven introns are short and similar in size, ranging in length from 77 to 110 bp in BCBL-R and from 78 to 96 bp in BC1. They also include several unusual nonconsensus donor or acceptor motifs associated with introns 4 and 5. The intact BCBL1 ORF-K15(P) protein encoded by the spliced cDNA clone proved to be nearly identical in amino acid sequence to that predicted from our genomic DNA sequence of BCBL-R (Fig. 5). In

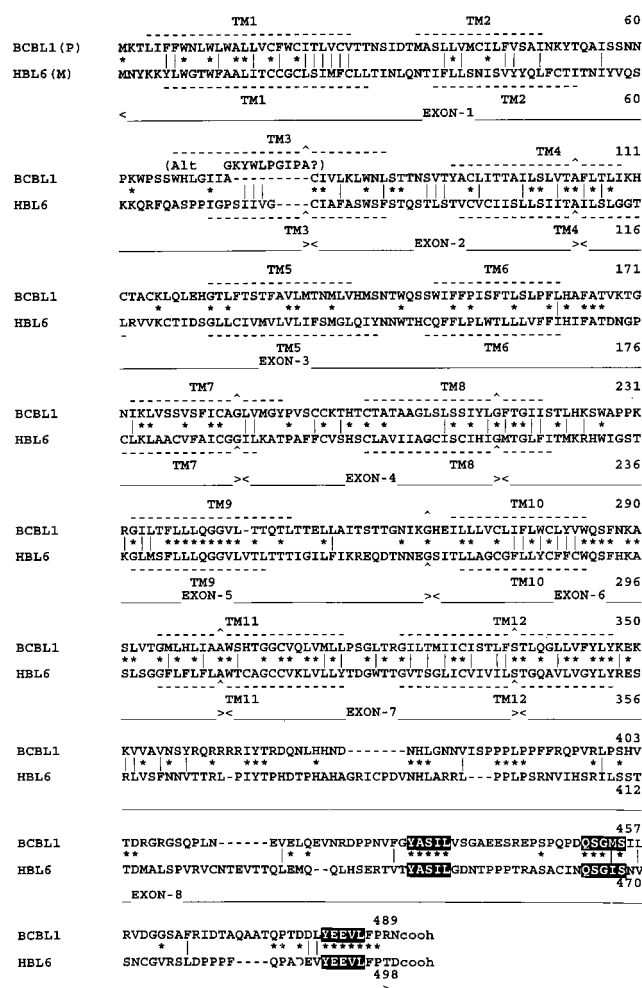


FIG. 5. Comparison of the complete predicted protein sequences of the ORF-K15(P) and ORF-K15(M) alleles of HHV8. Each protein consists of eight matching exons and 12 TM domains (TM1 to TM12) with an extended C-terminal cytoplasmic domain. The ORF-K15(P) allele coding region occupies nucleotide positions 1578 to 3652 and is oriented leftward in the HHV8(BCBL-R) and HHV8(BCBL1) DNA sequences (GenBank accession no. U85269). The ORF-K15(M) allele coding region occupies nucleotide positions 134671 to 136771 and is oriented leftward in the HHV8(BC1) DNA sequence (GenBank accession no. U75698) but has an additional A inserted after position 136610 in our HBL6 version. The 12 presumed TM domains are denoted by the broken overlines, and intron-exon junctions are indicated by carets. Amino acid identities are signified by asterisks, and similarities are shown by vertical lines. The potential protein tyrosine kinase binding motifs YASL and YEEVL plus a conserved QSG(M/I)S motif in the C-terminal cytoplasmic domain in exon 8 are highlighted.

addition, except for an apparent reading frame discrepancy in exon 1 of the sequence, our intact HBL6 cDNA clone encodes an ORF-K15(M) protein identical to that predicted from BC1 genomic DNA. Sequencing of the appropriate portions of HBL6, TKS1, and ASM72 genomic DNAs confirmed that they too contain an additional base pair in exon 1 near the N terminus of the protein relative to the database sequence for HHV8(BC1). This change extends the intact M allele ORF-K15 reading frame to an ATG initiator codon (at position 136771 in BC1) exactly equivalent to that in the P allele ORF-K15 genes in BCBL-R (position 3652) and BCBL1 (Fig. 5). The intact P and M versions of the ORF-K15 protein display an overall amino acid identity of only 33% with 50% similarity.

Both the P and M allele proteins contain predicted 142- to 145-amino-acid C-terminal cytoplasmic tail domains in exon 8,

which include conserved YXXL-like tyrosine kinase signalling motifs (e.g., YASI and YEEV) that would be expected to be targets for and to interact with SH2 domain protein kinases. They also have several possible Pro-rich SH3 domain-like motifs (SPPPLPP, PPLPS, PPPFQP, and TPPPT) and a conserved QSG(M/I)S motif. Curiously, this organization of probable C-terminal tail signalling motifs contrasts with the presence of a functional YXXLN6YXXL-containing immunoreceptor tyrosine-based activation motif (ITAM) signal plus a single YEEA motif and several SH3 domains in the 140-amino-acid N-terminal cytoplasmic tail instead in the EBV LMP2A protein. However, considering that there is no residual amino acid homology between them, the organizational features of the core integral membrane sections of the ORF-K15 and LMP2 genes display intriguing parallels, with TM domains 3, 4, 7, 8, and 12, all spanning introns, and the relative sizes of the first five coding exons being very similar (Fig. 4). Only the apparent fusion of exons 6 and 7 and the accompanying loss of the intron in TM domain 11 represent a significant change in the organization of this region in EBV LMP2 compared to the two HHV8 ORF-K15 alleles.

Overall, although the originally defined K15 ORF in HHV8 BC1 represents only exon 3 of the complete protein, the extended multiexon spliced genes for ORF-K15(P) and ORF-K15(M) occupy close to 2,100 bp each and essentially fill up the entire orphan region at the extreme RHS of the genomes between ORF75 and the TTR in both BCBL-R and HBL6/BC1. Our data have not permitted the identification of either the actual 5' or 3' ends of the ORF-K15 mRNAs, although we anticipate that as in LMP2 in EBV, the poly(A) signal is probably 3' coterminal with that of the downstream ORF75 gene. Interestingly, a proximal TATATAA box motif and probable associated lytic-cycle promoter elements are present just upstream of ORF-K15(M) exon-1 in BC1 DNA, but this is not the case in the ORF-K15(P) version, implying that there are likely to be significant differences in the transcriptional control of the two alleles.

The unique-sequence-TTR junction in the P form of the RHS end differs between the B/D and A/C subtypes. To further evaluate the level of heterogeneity found within the two forms of the extreme RHS of HHV8 genomes, we also analyzed another PCR locus representing the 850-bp unique-sequence-TTR boundary immediately upstream and to the right of the ORF-K15(P) genes from several genomes representing each of the major ORF-K1 subtypes. The results revealed two principal alternative PCR products that differed in size by approximately 45 bp, with the shorter forms being detected only in genomes with subtype B and D ORF-K1 genes (data not shown). As expected, this PCR primer pair failed to give any products with M subtype genomes.

Sequence data (not shown) for the U/TTR locus confirmed that all African KS genomes of the B/P type fell into two groups that could be readily distinguished from one another and from all A/P and C/P subtypes (for a summary, see Fig. 9). In particular, 431KAP, OKS4, RKS1, RKS3, and RKS4 all have either a 53-bp or a 55-bp deletion across the U/TTR boundary, as well as a nearby 10-bp insertion in intron 1 and up to 10 other single-nucleotide changes. We refer to these as B1 and B2 variants (see Fig. 9). In contrast, OKS3, RKS2, RKS5, and JKS15 form a very different subset, which we refer to as A/B (see Fig. 9) because they do not have the deletions and differ only slightly from the A/C pattern. The subtype D1 genome from TKS10 was similar to the deleted B1 and B2 patterns but included a second 15-bp insertion in the center of intron 1, as well as the 10-bp insertion and the 53-bp U/TTR junction deletion. Details of the upstream ORF-K15 regions

will be presented elsewhere. However, except for possible alternatively spliced forms of exon 1, there is little effect of any of these subtype B- and D-specific changes on the structure or amino acid sequence of the ORF-K15(P) protein.

In contrast to the B/P and D/P subtypes, the nine A/P and C/P subtypes sequenced showed minimal variations in U/TTR; all four of the subtype A1 or A4 genomes examined are identical to BCBL-R over the entire 849 bp, whereas the four subtype C3 or C4 genomes tested and the A3 variant (BCBL1) differ by only two to four nucleotides, none of which affects the ORF-K15 coding region. Note that the U/TTR boundaries in all of the P allele genomes analyzed are less than 300 bp upstream from the presumed ORF-K15(P) ATG initiator codon, whereas the upstream ORF-K15(M) region in BC1 contains an additional 3 to 4 kb of complex (G+A)-rich repeat sequence before the U/TTR boundary (31).

The M allele of ORF-K15 is linked to a specific M subtype of the adjacent ORF75 gene. To try to understand the differences noted previously between ASM72 and HBL6/BC1 compared to subtype A and B genomes within the ORF75 coding region, we PCR sequenced an expanded 854-bp region (ORF75-E) from 58 of the HHV8 genomes in our ORF-K1 set. The results revealed that even in the adjacent downstream ORF75 coding region (38), the influence of the M allele, versus that of the P allele, was still readily detectable. A chart showing the positions of all 20 of the nucleotide polymorphisms detected within ORF75-E and our interpretation of subgroup patterns is presented in Fig. 6. The data revealed that 25 of the 28 subtype A and C genomes having the P type of RHS were identical except for a single-nucleotide change in both subtype A3 genomes (BCBL1 and BKS14) and another in one of the subtype C2 samples (EKS1). Among the 14 subtype B genomes tested, two distinct patterns emerged. Six gave a characteristic B pattern, which differed from that of all of the A-plus-C versions at up to seven positions. The other eight samples (plus OKS3) were identical to the A/C pattern, except for 159-G in JKS20 and 563-G in the three American B samples. The three subtype D1 and D2 ORF75-E sequences were also unique and differed at six positions from the A/C pattern, as well as at seven positions from the B pattern. Remarkably, among genomes with the M type of RHS region, 14 of the 16 examples tested, representing a variety of C1, C2, C3, C5, A1, and A2 ORF-K1 genomes, were all identical to one another and to HBL6/BC1, except at one position. However, they also differed from the B/P subgroup ORF75-E pattern at 9 positions, from the D/P pattern at 13 positions, and from the A/P and C/P subtype RHS genomes at 14 positions. The two B/M genomes OKS7 and OKS8 from Florida were the only exceptions found in which M allele ORF-K15 genes were associated with P subtype ORF75-E patterns.

These results show that all subtype A and C genomes with the P form of RHS are virtually identical in both the ORF75-E block and the ORF-K14.1 junction regions that lie immediately to the left (downstream) of ORF-K15, although subtypes B and D both differ from them significantly (by 0.7 to 0.8% within the ORF75-E locus). On the other hand, both loci of all clade A or C genomes that have the M form of RHS are also identical to one another at both of these sites but give patterns that are quite distinct from those of the P type (by up to 1.5%). Therefore, all but 2 of the more than 50 genomes examined showed complete linkage between their P or M patterns for ORF75-E and ORF-K14.1 relative to their P or M alleles of ORF-K15. Overall, we conclude that HHV8 genomes with the ASM72 or HBL6/BC1 type of RHS are all chimeras containing an M variant of the RHS end, which includes very different ORF-K15 alleles and U/TTR junction regions, plus associated

ORFK1 BASED CLASS	ISOLATE NAME	POLYMORPHISMS (20 / 850 bp)																				ORF75 CLASS
		-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
		1	1	0	0	0	0	0	0	0	0	0	1	1	1	2	4	4	5	5	5	
	(* = M Allele at RHS)	3	2	2	2	1	0	0	1	2	6	7	5	6	9	3	1	6	2	6	9	
		7	5	9	3	5	4	9	8	0	3	5	9	2	2	7	7	2	8	3	1	
A1	BCBL-R	T	G	C	C	T	A	T	G	C	A	T	A	C	G	G	C	A	G	A	C	
A1	C282	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
A1	AKS1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
A1	AKS2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
A1	BKS11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
A1	SKS8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
A1'	BCP1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
A1'	BKS8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
A1'	ZKS5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
A1'	ZKS7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
A4	BCBL-B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
A4	BKS1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
A4	BKS10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
A4	ZKS1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
A4	OKS9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
A3	BCBL-1	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
A3	BKS14	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
C3	KS-F	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
C3	BC2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
C3	BC3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
C3	JSC1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
C3	BKS15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
C4	SKS1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
C2	SKS2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
C2	SKS6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
C2	SKS7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
C2	EKS1/2	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-		
A2 *	HBL6/BC1	C	-	-	T	-	C	A	-	T	G	C	G	-	A	A	-	G	T	G	-	
A1 *	BKS16	C	-	-	T	-	C	A	-	T	G	C	G	-	A	A	-	G	T	G	-	
C5 *	SKS9	C	-	-	T	-	C	A	-	T	G	C	G	-	A	A	-	G	T	G	-	
C3 *	BKS12	C	-	-	T	-	C	A	-	T	G	C	G	-	A	A	-	G	T	G	-	
C3' *	TKS1	C	-	-	T	-	C	A	-	T	G	C	G	-	A	A	-	G	T	G	-	
C3' *	TKS2	C	-	-	T	-	C	A	-	T	G	C	G	-	A	A	-	G	T	G	-	
C3' *	TKS3	C	-	-	T	-	C	A	-	T	G	C	G	-	A	A	-	G	T	G	-	
C3' *	TKS5	C	-	-	T	-	C	A	-	T	G	C	G	-	A	A	-	G	T	G	-	
C3' *	TKS7	C	-	-	T	-	C	A	-	T	G	C	G	-	A	A	-	G	T	G	-	
C3' *	TKS9	C	-	-	T	-	C	A	-	T	G	C	G	-	A	A	-	G	T	G	-	
C2 *	TKS11	C	-	-	T	-	C	A	-	T	G	C	G	-	A	A	-	G	T	G	-	
C2 *	SKS3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
C1 *	ASM72	C	-	-	T	-	C	A	-	T	G	C	G	-	A	A	-	G	T	G	-	
C1 *	BKS13	C	-	-	T	-	C	A	-	T	G	C	G	-	A	A	-	G	T	G	-	
B	431KAP	-	-	-	-	-	C	-	-	-	-	-	G	C	G	-	-	T	G	-	G	-
B	OKS4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
B	ST2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
B	RKS1	-	-	-	-	-	C	-	-	-	-	-	G	C	G	-	-	T	G	-	G	-
B	RKS3	-	-	-	-	-	C	-	-	-	-	-	G	C	G	-	-	T	G	-	G	-
B	RKS4	-	-	-	-	-	C	-	-	-	-	-	G	C	G	-	-	T	G	-	G	-
A5	OKS3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
B	ST1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
B	RKS2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
B	RKS5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
B	JKS20	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
B	JKS15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
B	* OKS7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
B	* OKS8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
D2	ZKS3	-	-	T	-	-	-	-	-	-	-	-	G	C	-	A	-	-	-	-	-	-
D2	ZKS4	-	-	T	-	-	-	-	-	-	-	-	G	C	-	A	-	-	-	-	-	-
D1	TKS10	-	-	T	-	-	-	-	-	-	-	-	G	C	-	A	-	-	-	-	-	-

FIG. 6. Comparison of polymorphic nucleotide patterns that identify four distinct subgroups of HHV8 genomes within the extended 854-bp ORF75-E block (map coordinate 0.96). Dashes indicate identities to the prototype A1 class sequence (BCBL-R) in the top line. The numbering system refers to positions (+ or -) relative to those used by Zong et al. (38) for ORF75 based on the ORF75 631-bp RDA fragment of Chang et al. (5). Positions -223, +1, and +631 are equivalent to positions 132875, 133098, and 133728 in the BC1 complete genome sequence (primers LGH2087 and LGH1704). ORF-K1 and ORF75-E subclass assignments are given in the far left-hand and far right-hand columns, respectively. Importantly, genomes with P allele ORF-K15 genes that have subtype A or C ORF-K1 genes (and some with subtype B ORF-K1 genes) cannot be distinguished here (=A/C class), whereas nearly all of the genomes tested with M allele ORF-K15 genes (denoted by asterisks) form a distinctive M subclass of ORF-75E patterns. Note that all of the genomes examined differed from the original 631-bp sequence of Chang et al. (5) by having 167-T, 442-C, and 496-T instead of 167-C, 442-T, and 496-C.

distinctive (but far less diverged) versions of the adjacent ORF-K14.1 and ORF75 genes also. Therefore, our original C prototype genome, ASM72, is now judged to be a probable C1/M recombinant whereas HBL6/BC1 represents an A2/M recombinant (see Fig. 9). Interestingly, 8 of the 18 genomes identified so far with the M type of RHS represent classic or

pretation is complicated by the following considerations. Firstly, five other genomes with subtype C2 ORF-K1 genes do not have a distinctive subtype C2 T0.7 pattern. Secondly, it does not seem very logical that the seven non-Taiwan-derived C3 patterns are all essentially indistinguishable from A1 patterns yet the novel T0.7 pattern displayed by 10 other subtype C genomes differ from the A1/C3 pattern at as many as 10 positions. Thirdly, we would also have to conclude that all other subtype C genomes (13 of 23) had recombined with subtype A genomes, although admittedly this must be the case anyway for 5 of them that show chimeric patterns (see below). Fourthly, and most compellingly, it is unlikely to be coincidental that all 10 of the genomes with this novel T0.7 pattern also have the M type of ORF-K15 gene. Overall, 10 of the 16 genomes with M genotype RHS ends that we tested gave this novel type of T0.7 pattern. Therefore we conclude, firstly, that (just as in ORF75-E) the A and C subtypes of P allele genomes are virtually indistinguishable and, secondly, that the highly diverged T0.7 pattern must reflect further leftward penetration of the M allele-associated RHS pattern.

Re-evaluation and extension of the ORF26 region data. Although we originally defined the major A, B, and C subtypes in the ORF-K1 region based on the presumption of linkage to the original assignments for the ORF26 and ORF75-C regions in our prototype strains (i.e., BCBL-R as A; 431KAP as B, and ASM72 as C), it has become evident that ORF26 data alone are insufficient for accurate discrimination between subtype B and C patterns, especially within the smaller 233-bp ORF26 block used in most previous studies. Partly because of the presumed need to revise some tentative assignments in Table 2 of Zong et al. (38) and partly because there were no subtype C3 genomes in our original analysis, we have re-evaluated a somewhat larger 330-bp ORF26 block at map position 0.35 in the genome after sequencing 60 of the HHV8 genomes included in the ORF-K1 analysis. A summary chart of the results and interpretations is presented in Fig. 8. The data include our re-analysis of BC2, which differs at two positions from the original published sequence of Cesarman et al. (3) for BC2, as well as our data for HBL6, which agrees with that for BC1 of Russo et al. (31) but not that of Cesarman et al. (3). We have also omitted the KSHV sequence of Chang et al. (5) because we suspect that the unusual base 1033-C, rather than 1033-T, in ORF26 and three nonconsensus bases in ORF75-E, which have not been found in any other samples, may all represent PCR errors introduced during representational difference analysis (RDA).

The new data revealed essentially identical patterns for all A1, A2, and A4 variants and the subtype D genomes but with a single C-to-A change in the two A3 variants at position 1103. In contrast, the prototype C1, C2, and C3 variants differed from the A pattern at 4, 3, and 5 positions, respectively, including changes at nucleotide positions 981-C, 1032-A, 1055-T, 1086-T, 1094-A, 1132-G, and 1139-C. Similarly, three distinct patterns emerged among known subtype B genomes, although one (B3) is identical to the C2 pattern, and all involve different arrangements of the same changes seen within the subtype C patterns at positions 981-C, 1032-A, 1132-G, and 1139-C. However, there is no evidence for any M-linked pattern in ORF26. Finally, we need to point out that the previously published ORF26 pattern for BC2 was originally assigned incorrectly as a B' type in Tables 2 and 3 of Zong et al. (38), but the corrected pattern is now recognized as the prototype for C3 variants seen also in BC3, BKS3, and JSC1.

Complexities and possible overlaps among some ORF26 patterns within the B and C subgroups. In the ORF-K1 analysis (37), only 1 of 10 African KS samples tested did not belong to subtype B, and that was the single example found of subtype

ORFK1 BASED CLASS	ISOLATE NAME	POLYMORPHISMS (10 / 330 bp)										ORF26 CLASS
		935	981	1032	1055	1086	1094	1103	1122	1132	1139	
		* = M Allele										
A1	BCBL-R	C	T	C	G	C	G	C	G	A	A	A
A1	C282	-	-	-	-	-	-	-	-	-	-	A
A1	AKS1	-	-	-	-	-	-	-	-	-	-	A
A1	AKS2	-	-	-	-	-	-	-	-	-	-	A
A1	AKS4	-	-	-	-	-	-	-	-	-	-	A
A1	BKS11	-	-	-	-	-	-	-	-	-	-	A
A1	SKS8	-	-	-	-	-	-	-	-	-	-	A
A1 *	BKS16	-	-	-	-	-	-	-	-	-	-	A
A1'	BCP1	-	-	-	-	-	-	-	-	-	-	A
A1'	BKS8	-	-	-	-	-	-	-	-	-	-	A
A1'	ZKS5	-	-	-	-	-	-	-	-	-	-	A
A1'	ZKS7	-	-	-	-	-	-	-	-	-	-	A
A4	BCBL-B	-	-	-	-	-	-	-	-	-	-	A
A4	BKS1	-	-	-	-	-	-	-	-	-	-	A
A4	BKS10	-	-	-	-	-	-	-	-	-	-	A
A4	OKS9	-	-	-	-	-	-	-	-	-	-	A
A4	ZKS1	-	-	-	-	-	-	-	-	-	-	A
A2 *	HBL6/BC1	-	-	-	-	-	-	-	-	-	-	A
A3	BCBL1	-	-	-	-	-	-	A	-	-	-	A3
A3	BKS14	-	-	-	-	-	-	A	-	-	-	A3
C3	KS-F	-	-	-	-	-	-	-	-	-	-	A
C3 *	BKS12	-	-	-	-	-	-	-	-	-	-	A
C2	SKS7	-	-	-	-	-	-	-	-	-	-	A
C2	EKS1/2	-	-	-	-	-	-	-	-	-	-	A
C3	BC2	-	C	A	T	-	-	-	-	G	C	C3
C3	BC3	-	C	A	T	-	-	-	-	G	C	C3
C3	JSC1	-	C	A	T	-	-	-	-	G	C	C3
C3	BKS15	-	C	A	T	-	-	-	-	G	C	C3
C3' *	TKS1	A	C	-	-	T	-	-	-	-	C	C2'
C3' *	TKS2	A	C	-	-	T	-	-	-	-	C	C2'
C3' *	TKS3	A	C	-	-	T	-	-	-	-	C	C2'
C3' *	TKS5	A	C	-	-	T	-	-	-	-	C	C2'
C3' *	TKS7	A	C	-	-	T	-	-	-	-	C	C2'
C3' *	TKS9	A	C	-	-	T	-	-	-	-	C	C2'
C5 *	SKS9	-	C	-	-	T	-	-	-	-	C	C2
C4	SKS1	-	C	-	-	T	-	-	-	-	C	C2
C2 *	TKS11	-	C	-	-	T	-	-	-	-	C	C2
C2	SKS2	-	C	-	-	T	-	-	-	-	C	C2
C2 *	SKS3	-	C	-	-	T	-	-	-	-	C	C2
C2	SKS6	-	C	-	-	T	-	-	-	-	C	C2
C1 *	ASM72	-	C	-	-	T	A	-	-	-	C	C1
C1 *	BKS13	-	C	-	-	T	-	-	-	-	C	C2
B	431KAP	-	C	-	-	-	-	-	-	G	C	B1
B	OKS4	-	C	-	-	-	-	-	-	G	C	B1
B	ST2/3	-	C	A	-	-	-	-	-	G	C	B2
B	RKS1	-	C	A	-	-	-	-	-	G	C	B2
B	RKS2	-	C	A	-	-	-	-	-	G	C	B2
B	RKS3	-	C	A	-	-	-	-	-	G	C	B2
B	RKS4	-	C	A	-	-	-	-	-	G	C	B2
B	JKS15	-	C	A	-	-	-	-	-	G	C	B2
B *	OKS7	-	C	A	-	-	-	-	-	G	C	B2
B *	OKS8	-	C	A	-	-	-	-	-	G	C	B2
A5	OKS3	-	C	-	-	T	-	-	-	-	C	B3/C2
B	ST1	-	C	-	-	T	-	-	-	-	C	B3/C2
B	RKS5	-	C	-	-	T	-	-	-	-	C	B3/C2
B	JKS20	-	C	-	-	T	-	-	-	-	C	B3/C2
B	ZKS6	-	C	-	-	T	-	-	T	-	C	B3/C2
D2	ZKS3	-	-	-	-	-	-	-	-	-	-	D/A
D2	ZKS4	-	-	-	-	-	-	-	-	-	-	D/A
D1	TKS10	-	-	-	-	-	-	-	-	-	-	D/A

FIG. 8. Comparison of polymorphic nucleotide patterns that identify several distinct subgroups of HHV8 genomes within the 330-bp ORF26 gene locus (map coordinate 0.35). Dashes indicate identities to the prototype A1 class sequence (BCBL-R) in the top line. The numbering system conforms to that used by Zong et al. (38) and introduced by Chang et al. (5) for the ORF26 RDA fragment. Positions 893 (=1) and 1222 (=330) in these PCR products are equivalent to positions 47193 and 47522 in the complete BC1 genomic sequence (primers LGH1701 and LGH1702). Overall, the ORF26 sequences resolve into four distinct patterns roughly equivalent to the ORF-K1 A, C3, C2, and B subtypes. However, the patterns for several of the B and C variants are either very similar or identical and the D patterns are indistinguishable from those of subtype A patterns. Therefore, the use of ORF26 patterns alone for classification is not recommended. Note that all of the genomes examined here differ from the original data of Chang et al. (5) by having 1033-T instead of 1033-C (genomic position 47333). All of the genomes examined here, including HBL6 and BC2, also differed from the data given for BC1 and BC2 by Cesarman et al. (3) by having 926-A, 989-C, and 1199-T.

A5 (OKS3). The ORF26 pattern for OKS3 is also distinctive, being unlike any of the other A patterns and different from the prototype C3 (BC2) and B (431KAP) patterns. However, the ORF26 sequence pattern for OKS3 matches those of TKS11 (C2) and BKS13 (C1), as well as our unusual ST1 chimeric African sample, which proved to have a subtype B ORF-K1 gene, although it clearly had an A/C subtype ORF75-E pattern. Furthermore, four other African KS samples reported by

Huang et al. (14), as well as four KS samples from Saudi Arabian renal transplant patients, have this type of ORF26 pattern (9). Subsequent sequencing of the ORF-K1 region of six Saudi Arabian samples with this type of ORF26 gene revealed four C2 patterns, as well as our prototype examples of the C4 and C5 patterns (37). Therefore, we now believe that the 981-C, 1086-T, and 1139-C pattern for ORF26, as found in TKS11, BKS13, SKS1, SKS2, SKS3, and SKS6, which all differ by only 1 bp (A-1094) from ASM72 (C1), probably represents the prototype C2 pattern. Since most C3 genomes have the distinctive 981-C, 1032-A, 1055-T, 1132-G, and 1139-C pattern for ORF26 (Fig. 8), we also have to conclude that the Taiwan clade TKS1 to TKS9 genomes, which are identical to C2 in ORF26 except for one additional distinctive change to 935-A, presumably all represent a novel C3'/C2'/M recombinant genome. Evidently, KS-F and EKS1 must both also represent chimeric C/A recombinant genomes based on discordant data for their ORF-K1 and ORF26 loci (see later).

The predominant subtype B patterns for ORF26 are either 981-C, 1132-G, and 1139-C (referred to as B1 variants) or 981-C, 1032-A, 1132-G, and 1139-C (referred to as B2 variants) (Fig. 8). However, four samples (ST1, RKS5, JKS20, and OKS3) all have the 981-C, 1086-T, and 1139-C pattern for ORF26, which could be a B3 subgroup but is identical to the C2 pattern (Fig. 8). Similarly, ZKS6 differs by only 1 bp (B3'). The B3 patterns may represent true B variants that happen to be identical to the C2 pattern, or alternatively, they might represent mosaic genomes. Overall, we conclude that there is considerable similarity among and overlap between the ORF26 patterns for several variants of the subtype B and C genomes. Furthermore, the subtype A and D patterns are not distinguishable either. Therefore, we caution that this ambiguity could lead to misinterpretations and incorrect strain assignments when judgements are based solely on ORF26 data, as has been the case in most previous analyses of HHV8 strain variability (8, 9, 14).

Correlations and linkage between multiple loci. An overall summary of the available data and subgroup assignments determined at each of up to seven distinct central and RHS loci across the genomes of 63 KS and PEL patient samples are compared with the LHS ORF-K1 data from Zong et al. (37) in Fig. 9. Ignoring the complexities introduced by the ORF-K15(M) RHS recombinants, a reasonably coherent picture has emerged showing linkage and cosegregation between subgroup patterns across the whole genome. In most of the 49 subtype A, C, and D genomes examined, the subgroup patterns of the ORF-K1, ORF26, T0.7, ORF75-E, and UPS75 loci are fully consistent across the entire DNA molecule. Obvious exceptions that are best interpreted as intertypic recombinants between different A and C variants include EKS1 and SKS7 (C2/A2), SKS3 (C2/A3), KS-F (C3/A1), and BKS12 (C3/A2), as well as the seven TKS1 to TKS9 genomes (C3'/C2'). Note that many of these also have an M allele pattern at the RHS. In addition, BCBL1, which may be an A3/C3 recombinant, has ORF26, T0.7, and ORF75-E patterns slightly different from those of all of the other subtype A samples (although in each case differing at only one position) but has a novel RHS U/TTR junction pattern (4-bp differences from all of the other subtype A and C samples tested), including a C3-like variant of the TTR. Our current interpretation of the overall structures of 13 clearly chimeric HHV8 genome types are illustrated and summarized in Fig. 10. Most of the intertypic and intratypic chimerism observed was associated with the nine distinct types of M allele-containing genomes, and it apparently reflects multiple sequential recombination events that generated these genomes, but four types of intertypic chimerism within P allele genomes are also illustrated.

NAME	TYPE	SOURCE	HETEROGENEOUS LOCI								OVERALL SUBTYPE
			K1	ORF26	T0.7	ORF75E	UPS75	K14.1	K15	U/TTR	
BCBL-R	B+	U	A1	A	A/C	A/C	A/C	A/C	P	A/C	A1/P
C282	K+	U	A1	A	A/C	A/C	A/C	A/C		A/C	A1/P
AKS1	K+	U	A1	A	A/C	A/C	A/C	A/C			A1/P
AKS2	K+	U	A1	A	A/C	A/C	A/C	(P)			A1/P
AKS4	K+	U	A1	A	A/C	A/C	A/C	(P)			A1/P
BKS11	K+	U	A1	A	A/C	A/C	(P)				A1/P
SKS8	KR	S	A1	A	A/C	A/C	(P)				A1/P
BKS16	K+	U	A1	A	A2	M	M	(M)			A1/A2/M
BCP1	B	U	A1'	A	A/C	A/C	(P)				A1/P
BKS8	K+	U	A1'	A	A/C	A/C	(P)				A1/P
ZKS5	K+	Z	A1'	A	A/C	A/C	(P)				A1/P
ZKS7	KR	Z	A1'	A	A/C	A/C	(P)				A1/P
BCBL-B	B+	U	A4	A	A/C	A/C	(P)		(P)	A/C	A4/P
BKS1	B+	U	A4	A	A2	A/C	(P)				A4/P
BKS10	K+	U	A4	A	A/C	A/C	(P)				A4/P
OKS9	K+	F	A4	A	A/C	A/C	(P)			A/C	A4/P
ZKS1	K+	Z	A4	A	A/C	A/C	(P)				A4/P
HLB6/BC1	B+	U	A2	A	A2	M	M	M	M	-	A2/M
WKS1	K+	U	A2	A			(M)				A2/M
BCBL1	B	U	A3	A3	A3	A3	A/C	A/C	P	A3	A3/P
BKS14	K+	U	A3	A3	A3	A3					A3/P
SKS9	KR	S	C5	C2	M	M		M			C5/M
KS-F	K+	G	C3	A	A/C	A/C					C3/A/P
BC2	B+	U	C3	C3	A/C	A/C	A/C	A/C	(P)	A/C	C3/P
BC3	B+	U	C3	C3	A/C	A/C	A/C	A/C	(P)	A/C	C3/P
JCS1	B+	U	C3	C3	A/C	A/C	(P)				C3/P
BKS3/4	KC	U	C3	A/C			(P)				C3/P
BKS12	KC	U	C3	A	A2	M	(M)				C3/A2/M
BKS15	K+	U	C3	C3	A/C	A/C	(P)				C3/P
TKS1	K+	T	C3'	C2'	M	M	M	(M)	(M)	-	C3'/C2/M
TKS2	KC	T	C3'	C2'	M	M	M	(M)	(M)	-	C3'/C2/M
TKS3	KC	T	C3'	C2'	M	M	M	(M)	(M)	-	C3'/C2/M
TKS5	KC	T	C3'	C2'	M	M	M	(M)	(M)	-	C3'/C2/M
TKS6	KC	T	C3'	C2'	M	M	M	(M)	(M)	-	C3'/C2/M
TKS7	KC	T	C3'	C2'	M	M	M	(M)	(M)	-	C3'/C2/M
TKS9	K+	T	C3'	C2'	M	M	M	(M)	(M)	-	C3'/C2/M
SKS1	KR	S	C4	C2	C4	A/C	(P)			A/C	C4/P
TKS11	KR	T	C2	C2	M	M	(M)	(M)			C2/M
SKS2	KR	S	C2	C2	A/C	A/C	A/C	(P)			C2/P
SKS3	KR	S	C2	C2	A3	M	(P)				C2/A3/M
SKS6	KR	S	C2	C2	A/C	A/C	(P)				C2/P
SKS7	KR	S	C2	A	A2	A/C	(P)				C2/A2/P
EKS1/2	KC	U	C2	A	A2	A/C	A/C	(P)			C2/A2/P
ASM72	K+	U	C1	C1	M	M	M	(M)		-	C1/M
BKS13	K+	U	C1	C2	M	M	M	(M)			C1/M
431KAP	KC	A	B	B1	B	B	B	B	P	B1	B/P
ST1	K+	A	B	B/C	A/C	A/C	B				B/A/P
SP2/3	K+	A	B	B2	B						B/P
OKS3	K+	A	A5	B/C	B1	A/C	(P)			A/B	A5/B/A/P
OKS4	K+	A	B	B1	B2	B	(P)				B/P
RKS1	K+	A	B	B2	B2	B	(P)				B/P
RKS2	K+	A	B	B2	B1	A/C	(P)			A/B	B/A/P
RKS3	K+	A	B	B2	B2	B	(P)				B/P
RKS4	K+	A	B	B2	B2	B	(P)				B/P
RKS5	K+	A	B	B/C	B1	A/C	(P)			A/B	B/A/P
ZKS6	K+	A	B	B/C	B1	B	(P)				B/P
JKS15	K+	U	B	B2	B3	A/C	B	(P)		A/B	B/A/P
JKS20	KC	A	B	B/C	B1	A/C	(P)				B/A/P
OKS7	K+	F	B	B/C	B3	A/C	(M)	(M)	-		B/A/M
OKS8	K+	F	B	B/C	B3	A/C	(M)	(M)	-		B/A/M
ZKS3	KC	P	D2	D	D	D	(P)				D2/P
ZKS4	KC	P	D2	D	D	D	(P)				D2/P
TKS10	KC	P	D1	D	D	D	(P)			D	D1/P

FIG. 9. Summary of the overall subclass assignments and patterns of linkage obtained at eight different loci across the HHV8 genome. The comparison includes all 63 HHV8 samples for which complete ORF-K1 sequence data (K1 column) were generated previously (37) and also includes updated ORF26 and UPS75 locus data relative to the original 12 genomes evaluated (38). ORF72 (vCYC-D) data are not included, and because there were only limited amounts of DNA available for some samples, sequences were not obtained at all loci in all cases (blanks). Type information in the column alongside the sample designations summarizes sample characteristics where + is HIV or AIDS associated, B is BCBL or PEL, K is KS, C is classic or endemic, and R is renal transplant. The source column indicates geographic locality (or origin), where A is Africa, F is Florida, G is Germany, P is Pacific Islands, S is Saudi Arabia, T is Taiwan, Z is New Zealand, and U is United States (other than Florida). Designations in parentheses in the K14.1 and K15 columns indicate data derived from PCR product size measurements only. All other entries are derived from PCR sequence data. Only P subtype information can be obtained at the U/TTR locus, and dashes in this column indicate where no PCR products were detected from M subtype genomes. Note that 9 samples are considered to be intertypic recombinants and 18 others are chimeras containing M allele sequences at the RHS (see Fig. 10).

The overall RHS genome assignments for the three D1 and D2 variants that form the Pacific ORF-K1 D subtype are of particular interest. Although the ORF-K1 patterns of the Hwalian sample from Taiwan (TKS10) and the two Polynesian samples (ZKS3 and ZKS4) are dramatically different (21% amino acid divergence), we had judged previously that there

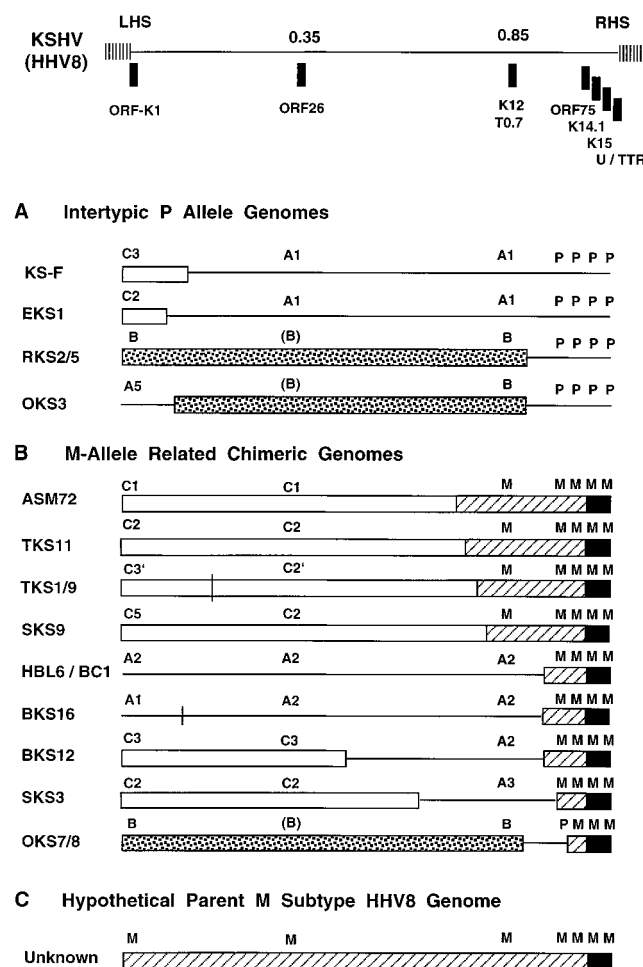


FIG. 10. Summary of the deduced overall genome structures of 13 different mosaic HHV8 genotypes that are interpreted to represent either intertypic recombinants or chimeras that have acquired ORF-K15(M) alleles and associated M-linked ORF75 and T0.7 genes. The relative genomic locations of the seven PCR loci involved in these studies are shown at the top. Key for recombinant genomes: solid horizontal lines, subtype A and RHS P allele sequences; open bars, subtype C sequences; dotted bars, subtype B sequences; hatched bars, M allele-linked sequences; solid bars, presumed exotic nonhuman ORF-K15(M) sequences. (A) Four examples of exclusively intertypic P allele genomes. Note that the EKS1 pattern is also found in SKS7 and that variations of the RKS2/5 pattern are also found in OKS3, ST1, JKS15, and JKS20. (B) Nine examples of chimeric genome patterns containing M allele RHS sequences, including both ORF-K15 (M) itself and adjacent M-linked genes. TKS1/9 and BKS16 are also considered to be intratypic recombinants independently of the M allele sequences, whereas BKS12, SKS3, and OKS7/8 are also intertypic recombinants. Note that the ASM72 pattern is also found in BKS13 and the TKS1/9 pattern is also found in TKS2, TKS3, TKS5, TKS6, and TKS7. (C) Model for the origin of M allele sequences from a hypothetical original intact M subtype parent genome. Note that this is itself interpreted to be a chimera containing a highly diverged ORF-K15(M) gene that was likely to have been derived initially by recombination between an ancient human M subtype HHV8 and an exotic nonhuman HHV8-like virus source.

were sufficient similarities to warrant their provisional assignment into a single novel D or Pacific subgroup (37). This view is now strengthened by the finding not only that these three samples were again distinguishable from all other subtype patterns at the T0.7, ORF75-E, UPS75, and U/TTR loci (Fig. 9) but that they were also almost indistinguishable from each other at these RHS loci, as well as in ORF26.

Complex chimerism among some subtype B genomes. Approximately half (7 of 12) of the subtype B KS samples from Africa that have been analyzed appear to have simple uncomplicated subtype B characteristics throughout their genomes.

However, the other five, as well as the three subtype B genomes from America, display a complex mosaicism that may reflect a history of both intertypic and intratypic recombination. Interestingly, all 10 of the samples obtained directly from Africa (as well as those from two recent African emigrants) have some features of subtype B genomes. This includes ST1, in which only the LHS is B, and OKS3, in which ORF-K1 is an A5 variant, but other parts of the genome (especially T0.7) have distinctive subtype B characteristics.

At the RHS of the mosaic subset of subtype B genomes, there is a common pattern for OKS3, RKS2, RKS5, JKS15, and JKS20, whereby they all have an A/C-like rather than B-like ORF75-E pattern, as well as a U/TTR pattern (referred to as A/B in Fig. 9) that is intermediate between the A/C and B patterns but much closer to A/C than to B. The two chimeric B/M genomes from Florida also have the A/C-like pattern for ORF75-E, although the U/TTR region in this case has been displaced by an M allele at the RHS. These findings are illustrated in a probably oversimplified form for the three genome structure prototypes exemplified by RKS2/5, OKS3, and OKS7/8 in Fig. 10. Subsequent analysis has also revealed that the LHS recombination boundary for both OKS7 and OKS8 shows a very distinct crossover point between the B pattern and the M-linked pattern within the N terminus of the ORF75 gene (data not shown). Importantly, all seven of these genomes have standard B pattern T0.7 genes but all except RKS2 also have the indeterminate B/C pattern ORF26 genes, which may or may not represent additional mosaicism toward the LHS. In addition, most of the subtype B patterns in constant-region loci (ORF26, T0.7, ORF75-E, and UPS75) can be divided relatively cleanly into at least three subpatterns (B1, B2, and B3 variants). However, there appears to be significant discordance among these subpatterns at different loci in individual genomes, suggesting the possibility of additional ancient intratypic mosaicism. Obviously, a great deal more analysis is required to define the overall evolutionary history of many of these subtype B genomes. However, based on the results of Kasolo et al. (15), who found that 15 of 15 HHV8-positive children in Zambia had subtype A5 ORF-K1 genomes in their peripheral lymphocytes, as well as from our analysis of two other A5/B pattern mosaic genomes from South Africa (1), this OKS3-like class of genomes may be quite prevalent and important in Africa.

DISCUSSION

The ORF-K1 genes from more than 60 HHV8 genomes analyzed previously have evolved into four major subtypes with distinctive ethnic and geographic associations that we believe are likely to have arisen during the migratory divergence of modern humans in Paleolithic times (37). Thus, the isolation of the precursor D subtype of HHV8 from a precursor B subtype was predicted to have occurred approximately 60,000 years ago as humans first spread from Africa to Southern Asia and the Pacific, whereas the derivation and subsequent divergence of the A and C branches probably correlated with later waves of migration into Europe and Northern Asia via the Middle East about 35,000 years ago. The nature of the biological selection pressure that has been driving the unusually high level of amino acid diversity displayed by the ORF-K1 membrane receptor signalling protein (18, 20, 28) is not understood, but obvious additional questions have arisen about whether or not the ORF-K1 protein subtypes also contribute to the different disease patterns observed in different regions of Africa compared to Mediterranean countries, etc. However, before this concept can be evaluated seriously, we considered it necessary to resolve whether other segments of the HHV8 genome dis-

play parallel strain heterogeneity patterns, as well as to determine whether previous hints that some genomes may be chimeric recombinants between subtypes represent additional complicating factors.

Therefore, we carried out a systematic analysis of sequence variability at three internal constant-region loci and at the extreme RHS of the genome in the same large set of samples that had been analyzed for ORF-K1 (37). The results of these studies at genomic map coordinates 0.35 (ORF26), 0.85 (T0.7/K12), and 0.96 (ORF75-E), as well as extensive evaluation of an apparent dichotomy in the DNA sequence patterns at the RHS of the genome (27), have led to three major conclusions. Firstly, consistent subtype A, B, C, and D patterns matching those in ORF-K1 are detectable throughout the more conserved internal regions of the genome, although the level of divergence is greatly diminished. Secondly, a very different pattern of strain subgrouping occurs at the extreme RHS of the genome, whereby two alternative highly diverged P and M subtypes or alleles of the ORF-K15 protein are essentially unlinked to the LHS subtype patterns. Thirdly, a combination of the presence of distinctive alleles of other adjacent genes linked to the P and M forms, together with evidence for simple recombination between some major subtypes, complicates the analysis of strain assignments in as many as 30% of all HHV8 genomes.

Origins of the ORF75 and T0.7 M-linked alleles. Our original interpretation (38) that the ORF75 sequence pattern present in both the ASM72 and HBL6/BC1 genomes represented the prototype subtype C pattern has had to be revised. Instead, this pattern clearly represents that of the M-linked allele of ORF75, which was present in 14 of the 16 genomes examined that have the M subtype of the ORF-K15 gene, and includes examples of C2, C3, C5, A1, and A2 variant ORF-K1 genomes. Furthermore, examination of a number of ORF-K15 (P) allele genomes having subtype C ORF-K1 genes revealed that the presumed true C pattern for ORF75-E cannot be resolved from that of most subtype A genomes, although the B and D patterns are distinctive. The level of nucleotide divergence of the subtype P[D] and P[B] ORF75-E loci from P[A/C] is 0.7% and 0.8% respectively, whereas the M allele ORF75-E pattern differs from P[A/C] by 1.4% and from P[B] by 1.0%. Therefore, it is tempting to argue that the M allele of ORF75 represents an older form that diverged from the P prototype before modern HHV8 genomes split into the A, B, C, and D subgroups. Similarly, at the T0.7 locus, all remaining M patterns (10 of the 18) are linked to both ORF75(M) and ORF-K15(M) genes, and again the M form is at least as far diverged from the P[A/C] and P[D] patterns as is the P[B] pattern.

There are two plausible explanations for these findings. Firstly, the acquisition of the ORF-K15(M) genes from some exotic but related primate HHV8-like virus increased the rate of evolutionary divergence of the adjacent genes in a prototype chimeric virus. A second alternative is that the exotically acquired ORF-K15(M) gene was originally associated with an evolutionarily older and presumably intact M subtype of HHV8 genome rather than with the more modern P subtype. Either way, we see virtually no sequence variation at all within ORF-K15(M) or ORF75(M) genes among the 18 genomes identified so far that carry the M alleles. Therefore, a single relatively recent event must have subsequently introduced this apparent relic from the RHS only of an ancient M subtype virus into a modern P subtype human virus. This event must have occurred at some point after the subtype C lineage diverged from the subtype A lineage but early enough that some descendent chimeric genomes could spread into both Europe and Asia, as well as into the Middle East. Therefore, the M

alleles of ORF75 and T0.7 probably represent part of a novel virus from an intermediate evolutionary stage of the modern human lineage dating back to between 100,000 and 300,000 years ago and that the subsequent recombination event that introduced it into a modern human P subtype genome occurred in the Middle East, perhaps between 25,000 and 35,000 years ago. Clearly, many of the chimeric viruses found represent even more recent transfer into other subtype A, B, and C genomes by further relatively rare homologous recombination events. Importantly, no M forms have been detected among African or Pacific samples and the survival and apparent spread of this chimeric form of the virus within many classical KS patients, as well as AIDS patients, may imply some form of selective advantage for the M allele of ORF-K15 over the P allele.

Specific features of M subtype recombinant genomes. The viruses found in all seven of the classic and AIDS-associated KS lesions examined from patients of Northern Chinese heritage in Taiwan (whose ancestors first emigrated no more than 400 years ago) represent a particularly interesting and complex clade. Although these seven genomes all differ from one another in the VR* loop and other variable loci within ORF-K1 (37), they all contain a distinctive subset of C3' ORF-K1 genes and a unique C2' variant of the ORF26 gene, together with both an apparently invariant M allele of the ORF-K15 gene and linked, almost invariant M subtypes of T0.7 and ORF75 (Fig. 9 and 10). Other than this clade, only 11 of the 56 genomes that we have examined from elsewhere around the world contained the ORF-K15(M) type of RHS. Note that in contrast, all five of the classic KS patients of Hwalien heritage that have been tested (whose ancestors are believed to have arrived in Taiwan from the Pacific Islands about 4,000 years ago) had D1/P subtype HHV8 genomes (33a, 37).

Irrespective of the origin of the linked M alleles of ORF75 and T0.7, all of the genomes examined that contained these genes are clearly chimeras that retain RHS segments only from the unknown exotic source virus (Fig. 10). The largest fragment present extends back beyond T0.7 at map position 0.85 but certainly not as far as ORF26 at 0.35. We have identified four distinct sets of genomes of this type that all have subtype C ORF-K1 genes, namely, (i) ASM72 and BKS13 (both C1/M), (ii) SKS9 (C5/M), (iii) TKS11 (C2/M), and (iv) all seven examples of the Taiwan TKS1-to-TKS9 clade (C3'/C2'/M). A smaller segment (not including the M allele of T0.7) was then presumably secondarily transferred to A2 and A1 variants exemplified by HBL6/BC1 and BKS16. Since the only other United States sample found that contains M sequences (BKS12) also has an A2-like T0.7 gene within a C3/A2/M chimeric genome, we presume that this also represents a secondary transfer recombination event. Finally, both the unusual Saudi Arabian C4/A3/M genome in SKS3 and the two Florida B/A/M recombinants (OKS7 and OKS8) also appear to represent alternative and less extensive secondary transfer lineages (Fig. 10). The two Florida AIDS KS samples came from patients of black Haitian and hispanic Mexican ethnicity, suggesting that they arose as a result of interactions between African and Hispanic cultures in Central America within the past few hundred years. Overall, in addition to the two Florida cases, we have identified only three other subtype B ORF-K1 genomes outside of Africa. Two were derived from African emigrants from Africa, one being a classic KS case in Baltimore (JKS20) and another an AIDS-associated case in New Zealand (ZKS6), whereas the third was from an African American with AIDS in Baltimore (JKS15), but none of these were M recombinants.

Lack of divergence between subtypes A and C at the RHS is paralleled within the constant regions of ORF-K1. Our inter-

pretation that all subtype A and C genomes with P allele RHS patterns are almost invariant within their T0.7, ORF75-E, and ORF-K15 genes was initially rather surprising considering that the A and C subtypes of ORF-K1 at the LHS of the genome differ by as much as 14% at the amino acid level. However, a re-evaluation of the distribution of sequence variations across relatively well-conserved segments of the ORF-K1 membrane signalling protein confirms the notion that they are, in fact, much more closely related than the overall level of variation indicates. As we pointed out previously (37), there are two distinct types of variation in ORF-K1, firstly, the relatively consistent and characteristic intertypic variations and, secondly, the seemingly random intratypic hypervariability displayed both between and within the different variants and clades of each subtype. Both levels of divergence appear to result from powerful biological selection pressures, but the latter are concentrated mostly within the two 40-amino-acid blocks (VR1 and VR2) within the extracellular domains and especially within the 23-amino-acid Cys-bridged VR* loop (37). In fact, most of the variations between subtypes A and C of ORF-K1 occur within these two hypervariable domains. In contrast, when only the four most conserved regions (totaling nearly 60% of the protein) from between amino acids 1 to 36, 105 to 147, 173 to 199, and 229 to 289 are considered, the prototype A and C subtypes differ by only 2.5% (six amino acid changes), whereas A and D still differ by 19% (32 of 164) and A and B differ by 27% (47 of 164). Therefore, we conclude that the hypervariability displayed within and between the A and C subtypes and variants predominantly represents a much more recent ongoing intratypic process (that has occurred even within the Taiwan C3' clade, for example). In contrast, the specific variations common to all members of a subtype obviously represent a more accurate reflection of older evolutionary divergence. Therefore, in this sense, even on the LHS, the much greater similarity between the constant regions of the subtype A and C proteins compared to subtype B and D proteins matches quite closely to the patterns observed at the RHS. A similar situation applies to the D1 and D2 subtypes, which are virtually identical at the ORF26, T0.7, ORF75, and UPS75 gene loci, despite showing 21% overall amino acid differences between them in ORF-K1.

Nature and origin of the two types of ORF-K15 genes. The M and P subtypes (or alleles) of the HHV8 ORF-K15 integral membrane protein genes, although they display less than 30% amino acid identity, have practically identical intron-exon splicing patterns in latent-state mRNA, and they show considerable overall structural and exon pattern resemblances (but no homology) to the LMP2 latency protein of EBV. All three proteins have 12 similarly spaced hydrophobic TM domains, and in fact, our initial prediction of the complete ORF-K15 protein structure, which has been confirmed by RT-PCR analysis, depended partly upon the presumption of a linear splicing pattern closely parallel to that of LMP2 (17). However, the ORF-K15 proteins contain two conserved likely SH2-binding tyrosine kinase signalling motifs and other proline-rich SH3-like binding domains within a large C-terminal cytoplasmic domain, whereas LMP2 has functional ITAM, SH2, and SH3 binding domains in an N-terminal cytoplasmic domain originating from across the other side of the TTRs (Fig. 4). Nevertheless, there appears to be little doubt, because of their equivalent genomic positions and orientation, as well as the similar sizes and splicing patterns of exons 2 to 7, that they are evolutionarily related genes.

There is no ORF-K15-like gene in HVS; however, several HHV8-like viruses have been described in Old World primates (7, 30) and the complete sequence of one of these, referred to

as rhesus rhadinovirus (RRV), has recently been presented (32). Although the presence or absence of an ORF-K15-like gene in RRV was not addressed by those authors, one can predict from the sequence at the extreme RHS that RRV does contain an equivalent highly spliced gene encoding a protein with a structure similar to that of HHV8 ORF-K15 but without any significant residual amino acid homology. Therefore, we conclude that even the exotic ORF-K15(M) gene must have originated within an HHV8-like gammaherpesvirus. It seems probable that the 33% amino acid identity between ORF-K15(M) and ORF-K15(P) represents a level of evolutionary divergence that is consistent with an origin for one of them (the M allele form) by recombination with a related virus from a great ape or other Old World primate species. In comparison, the LMP2A genes of human EBV and the equivalent gamma-1 class virus from baboons (*herpesvirus papio*) still display 50% amino acid identity (10). However, the possibility that there was once more than one HHV8-like virus with this level of divergence within the human lineage cannot be ruled out.

The problem of the origin of the two very distinctive subtype A and B EBNA2 genes in EBV (65% amino acid identity compared to 55% identity between EBV and *herpesvirus papio*) (6) and of the three even further diverged alternative STP-like genes within *herpesvirus saimiri* provides similar scenarios (19, 21). Do these alleles all represent divergent forms of key genes that suddenly evolved very rapidly with all intermediate forms being lost as each subtype of the virus occupied a new biological niche, or do they instead (as suggested here) represent relics of older forms of the virus or of related viral species that may now persist only as small pieces of their original genomes by virtue of rare recombination events with more modern forms? Our DNA sequence data from multiple samples of both the P and M subtypes of HHV8 show that neither of the two ORF-K15 alleles displays any more nucleotide variation than do the adjacent conserved genes of this virus; this appears to strongly support the recombination model. On the other hand, the hypervariability in ORF-K1 obviously represents an example of a gene that is currently undergoing very rapid evolution within the modern form of HHV8. The equivalent ORF-R1 gene in RRV has also diverged so far as to have very low residual homology to ORF-K1 of HHV8 (32). While we obviously favor the interpretation that a small number of very rare recombination events gave rise to all of the chimeric P and M class genomes described here, as well as to most of the intratypic and intertypic chimeras detected, the possibility that more-complex mosaics, such as those seen in many of the subtype B genomes, represent coordinate evolutionary drift at multiple loci cannot be ruled out.

ACKNOWLEDGMENTS

These studies were funded by Public Service grant R01 CA73585 to G.S.H. from the National Cancer Institute, N.I.H. L.J.P. was supported by a graduate student stipend from the Biochemistry, Cell and Molecular Biology Training Program at the Johns Hopkins School of Medicine (2T32-GM07445).

We thank Margit Lucskay for technical assistance and Sarah Heagans for help with preparation of the manuscript. We thank our colleagues K. Foreman, B. Nickoloff, and S. Alkan (SKS series); P. Browning, P. Rady, and S. Tying (BKS); K. Powell and M. Croxson (ZKS); I.-J. Su (TKS); and C. Rabkin (RKS), as well as A. Blauvelt and the AIDS Malignancy Bank (OKS) for providing various KS DNA or frozen and paraffin block samples.

ADDENDUM

We are aware that A. Davison (6a) has made similar independent predictions about the ORF-K15 structure, with sup-

porting evidence from cDNA analysis carried out by Glenn et al. (12a).

REFERENCES

- Alagiozoglou, L., J. C. Zong, and G. S. Hayward. Unpublished data.
- Beral, V., T. A. Peterman, R. C. Berkelman, and H. W. Jaffe. 1990. Kaposi's sarcoma among persons with AIDS: a sexually transmitted infection? *Lancet* 335:123-128.
- Cannon, J. S., A. C. Hawkins, C. A. Griffin, Q. Tao, M. Borowitz, G. S. Hayward, and R. F. Ambinder. Characterization of a new EBV⁺/HHV8⁺ primary effusion lymphoma-derived cell line. *J. Infect. Dis.*, in press.
- Cesarman, E., P. S. Moore, P. H. Rao, G. Inghirami, D. M. Knowles, and Y. Chang. 1995. In vitro establishment and characterization of two AIDS-related lymphoma cell lines containing Kaposi's sarcoma-associated herpesvirus-like (KSHV) DNA sequences. *Blood* 86:2708-2714.
- Cesarman, E., R. G. Nador, F. Bai, R. A. Bohenzky, J. J. Russo, P. S. Moore, Y. Chang, and D. M. Knowles. 1996. Kaposi's sarcoma-associated herpesvirus contains G protein-coupled receptor and cyclin D homologs which are expressed in Kaposi's sarcoma and malignant lymphoma. *J. Virol.* 70:8218-8223.
- Chang, Y., E. Cesarman, M. S. Pessin, F. Lee, J. Culpepper, D. M. Knowles, and P. S. Moore. 1994. Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science* 266:1865-1869.
- Dambaugh, T., K. Hennessey, L. Channankit, and E. Kieff. 1984. U2 region of Epstein-Barr virus DNA may encode Epstein-Barr virus nuclear antigen 2. *Proc. Natl. Acad. Sci. USA* 81:7632-7636.
- Davison, A., et al. Personal communication.
- Desrosiers, R. C., V. G. Sasseville, S. C. Czajak, X. Zhang, A. A. Lackner, and J. U. Jung. 1997. A herpesvirus of rhesus monkeys related to the human Kaposi's sarcoma-associated herpesvirus. *J. Virol.* 71:9764-9769.
- Di Alberti, L., A. Piattelli, L. Artese, G. Favia, S. Patel, N. Saunders, S. R. Porter, C. M. Scully, S.-L. Ngul, and C.-G. Teo. 1997. Human herpesvirus 8 variants in sarcoid tissues. *Lancet* 350:1655-1659.
- Foreman, K. E., S. Alkan, A. E. Krueger, J. R. Panella, L. J. Swinnen, and B. J. Nickoloff. 1998. Geographically distinct HHV-8 DNA sequences in Saudi Arabian iatrogenic Kaposi's sarcoma lesions. *Am. J. Pathol.* 153:1001-1004.
- Franken, M., B. Annis, A. N. Ali, and F. Wang. 1995. 5' coding and regulatory region sequence divergence with conserved function of the Epstein-Barr virus LMP2A homology in herpesvirus papio. *J. Virol.* 69:8011-8019.
- Gao, S.-J., L. Kingsley, D. R. Hoover, T. J. Spira, C. R. Rinaldo, A. Saah, J. Phair, R. Detels, P. Parry, Y. Chang, and P. S. Moore. 1996. Seroconversion to antibodies against Kaposi's sarcoma-associated herpesvirus-related latent nuclear antigens before the development of Kaposi's sarcoma. *N. Engl. J. Med.* 335:233-241.
- Gao, S.-J., L. Kingsley, M. Li, W. Zheng, C. Parravicini, J. Ziegler, R. Newton, C. R. Rinaldo, A. Saah, J. Phair, R. Detels, Y. Chang, and P. S. Moore. 1996. KSHV antibodies among Americans, Italians and Ugandans with and without Kaposi's sarcoma. *Nat. Med.* 2:925-928.
- Glenn, M., L. Rainbow, F. Auradé, A. Davison, and T. F. Schulz. 1999. Identification of a spliced gene from Kaposi's sarcoma-associated herpesvirus encoding a protein with similarities to latent membrane proteins 1 and 2A of Epstein-Barr virus. *J. Virol.* 73:6953-6963.
- Harwood, A. R., D. Osoba, S. L. Hofstadter, M. B. Goldstein, C. Cardella, M. J. Holecck, R. Kunyetz, and R. A. Giammarco. 1979. Kaposi's sarcoma in recipients of renal transplants. *Am. J. Med.* 67:759-765.
- Huang, Y. Q., J. J. Li, M. H. Kaplan, B. Poiesz, E. Katabira, W. C. Zhang, D. Feiner, and A. E. Friedman-Kien. 1995. Human herpesvirus-like nucleic acid in various forms of Kaposi's sarcoma. *Lancet* 345:759-761.
- Kasolo, F. C., M. Monze, N. Obel, R. A. Anderson, C. French, and U. A. Gompels. 1998. Sequence analyses of human herpesvirus-8 strains from both African human immunodeficiency virus-negative and -positive childhood endemic Kaposi's sarcomas show a close relationship with strains identified in febrile children and high variation in the K1 glycoprotein. *J. Gen. Virol.* 79:3055-3065.
- Kedes, D. H., E. Oberskalski, M. Busch, R. Kohn, J. Flood, and D. Ganem. 1996. The seroepidemiology of human herpesvirus 8 (Kaposi's sarcoma-associated herpesvirus): distribution of infection in KS risk groups and evidence for sexual transmission. *Nat. Med.* 2:918-924.
- Laux, G., M. Perricaudet, and P. J. Farrell. 1988. A spliced Epstein-Barr virus gene expressed in immortalized lymphocytes is created by circularization of the linear viral genome. *EMBO J.* 7:769-774.
- Lee, H., J. Guo, M. Li, J.-K. Choi, M. DeMaria, M. Rosenzweig, and J. U. Jung. 1998. Identification of an immunoreceptor tyrosine-based activation motif of K1 transforming protein of Kaposi's sarcoma-associated herpesvirus. *Mol. Cell. Biol.* 18:5219-5228.
- Lee, H., J. J. Trimble, D.-W. Yoon, D. Regier, R. C. Desrosiers, and J. U. Jung. 1997. Genetic variation of herpesvirus saimiri subgroup A transforming protein and its association with cellular src. *J. Virol.* 71:3817-3825.
- Lee, H., R. Veazey, K. Williams, M. Li, J. Guo, F. Neipel, B. Fleckenstein, A. Lackner, R. C. Desrosiers, and J. U. Jung. 1998. Deregulation of cell growth by the K1 gene of Kaposi's sarcoma-associated herpesvirus. *Nat. Med.* 4:435-440.
- Medveczky, P., E. Szomolanyi, R. C. Desrosiers, and C. Mulder. 1984. Classification of herpesvirus saimiri into three groups based on extreme variation in a DNA region required for oncogenicity. *J. Virol.* 52:938-944.
- Moore, P. S., C. Boshoff, R. A. Weiss, and Y. Chang. 1996. Molecular mimicry of human cytokine and cytokine response pathway genes by KSHV. *Science* 274:1739-1744.
- Moore, P. S., S.-J. Gao, G. Dominguez, E. Cesarman, O. Lungu, D. Knowles, R. Garber, P. E. Pellett, D. J. McGeoch, and Y. Chang. 1996. Primary characterization of a herpesvirus agent associated with Kaposi's sarcoma. *J. Virol.* 70:549-558.
- Neipel, F., J.-C. Albrecht, and B. Fleckenstein. 1997. Cell-homologous genes in the Kaposi's sarcoma-associated rhadinovirus human herpesvirus 8: determinants of its pathogenicity? *J. Virol.* 71:4187-4192.
- Nicholas, J., V. Ruvalo, J. Zong, D. Ciuffo, H.-G. Guo, M. S. Reitz, and G. S. Hayward. 1997. A single 13-kilobase divergent locus in Kaposi sarcoma-associated herpesvirus (human herpesvirus 8) genome contains nine open reading frames that are homologous to or related to cellular proteins. *J. Virol.* 71:1963-1974.
- Nicholas, J., V. R. Ruvalo, W. H. Burns, G. Sandford, X. Wan, D. Ciuffo, S. B. Hendrickson, H.-G. Guo, G. S. Hayward, and M. S. Reitz. 1997. Kaposi's sarcoma-associated human herpesvirus-8 encodes homologues of macrophage inflammatory protein-1 and interleukin-6. *Nat. Med.* 3:287-292.
- Nicholas, J., J.-C. Zong, D. J. Alcendor, D. M. Ciuffo, L. J. Poole, R. T. Sarisky, C. J. Chiou, X. Zhang, X. Wan, H.-G. Guo, M. S. Reitz, and G. S. Hayward. 1998. Novel organizational features, captured cellular genes and strain variability within the genome of KSHV/HHV8. *J. Natl. Cancer Inst. Monogr.* 23:79-88.
- Poole, L. J., D. M. Ciuffo, B. Chandran, and G. S. Hayward. Identification and expression of the immunoglobulin receptor-like glycoprotein encoded by the hypervariable ORF-K1 transforming gene of KSHV/HHV8. Submitted for publication.
- Qunibi, W. Y., Y. Barri, O. Alfurayh, K. Almehari, B. Khan, S. Taher, and K. Sheth. 1993. Kaposi's sarcoma in renal transplant recipients: a report on 26 cases from a single institution. *Trans. Proc.* 25:1402-1405.
- Rose, T. M., K. B. Strand, E. R. Schultz, J. G. W. Rankin, M. E. Thoulous, C.-C. Tsai, and M. L. Bosch. 1997. Identification of two homologs of the Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) in retroperitoneal fibromatosis of different macaque species. *J. Virol.* 71:4138-4144.
- Russo, J. J., R. A. Bohenzky, M.-C. Chien, J. Chen, M. Yan, D. Maddalena, J. P. Parry, D. Peruzzi, I. S. Edelman, Y. Chang, and P. S. Moore. 1996. Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus (HHV8). *Proc. Natl. Acad. Sci. USA* 93:14862-14867.
- Searles, R. P., E. P. Bergquam, M. K. Axthelm, and S. W. Wong. 1999. Sequence and genomic analysis of rhesus macaque rhadinovirus with similarity to Kaposi's sarcoma-associated herpesvirus/human herpesvirus 8. *J. Virol.* 73:3040-3053.
- Staskus, K. A., W. Zhong, K. Gebhard, B. Herndier, H. Wang, R. Renne, J. Benke, J. Pudney, D. J. Anderson, D. Ganem, and A. T. Haase. 1997. Kaposi's sarcoma-associated herpesvirus gene expression in endothelial (spindle) tumor cells. *J. Virol.* 71:715-719.
- Su, I. J. Personal communication.
- Sun, R., S.-F. Lin, L. Gradoville, and G. Miller. 1996. Polyadenylated nuclear RNA encoded by Kaposi sarcoma-associated herpesvirus. *Proc. Natl. Acad. Sci. USA* 93:11883-11888.
- Whitby, D., M. R. Howard, M. Tenant-Flowers, N. S. Brink, A. Copas, C. Boshoff, T. Hatziioannou, R. E. Suggett, D. M. Aldam, A. S. Denton, et al. 1995. Detection of Kaposi sarcoma associated herpesvirus in peripheral blood of HIV-infected individuals and progression to Kaposi's sarcoma. *Lancet* 346:799-802.
- Zhong, W., H. Wang, B. Herndier, and D. Ganem. 1996. Restricted expression of Kaposi sarcoma-associated herpesvirus (human herpesvirus 8) genes in Kaposi sarcoma. *Proc. Natl. Acad. Sci. USA* 93:6641-6646.
- Zong, J.-C., D. M. Ciuffo, D. J. Alcendor, X. Wan, J. Nicholas, P. J. Browning, P. L. Rady, S. K. Tyring, J. M. Orenstein, C. S. Rabkin, I.-J. Su, K. F. Powell, M. Croxson, K. E. Foreman, B. J. Nickoloff, S. Alkan, and G. S. Hayward. 1999. High-level variability in the ORF-K1 membrane protein gene at the left end of the Kaposi's sarcoma-associated herpesvirus genome defines four major virus subtypes and multiple variants or clades in different human populations. *J. Virol.* 73:4156-4170.
- Zong, J.-C., C. Metroka, M. S. Reitz, J. Nicholas, and G. S. Hayward. 1997. Strain variability among Kaposi sarcoma-associated herpesvirus (human herpesvirus 8) genomes: evidence that a large cohort of United States AIDS patients may have been infected by a single common isolate. *J. Virol.* 71:2505-2511.